

THIS WEEK

EDITORIALS

WORLD VIEW Arab revolutions positive for science — for now **p.131**

FOOD OF LOVE Frisky flies get picked off by hungry bats **p.132**

HUMANITIES Language and art around for longer than we thought **p.133**



Who calls the shots?

US law-makers need to encourage research on firearms-related violence so that gun laws can be based on facts rather than ideology.

The shooting spree in a Colorado cinema last month that killed 12 people and injured 58 has provoked many questions and much soul-searching. Some reports have even suggested that because the perpetrator, James Holmes, was until just weeks earlier a graduate student supported by a US National Institutes of Health (NIH) training grant, the biomedical agency is somehow implicated.

The insinuation is ludicrous, but the attention it received speaks volumes about the political reluctance in the United States to address the laws that made it possible for Holmes to obtain his arsenal of firearms. In this climate, discussions of the multiple murders sounded all too often like descriptions of the random and inevitable carnage caused by a tornado or an earthquake.

Natural disasters, truly unavoidable events, can be combated with science. The US Geological Survey, for instance, has some 250 employees dedicated to the assessment of earthquake hazards.

There is no such US government effort for research on firearms — the National Rifle Association has helped see to that. The lobby group that represents gun owners began to squash scientific efforts in 1996, when, using proxies in Congress, it shut down a fledgling, US\$2.6-million gun-violence research effort by the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia.

Among other things, that work revealed that people living in homes where there was a gun faced a 2.7-fold greater risk of homicide (A. L. Kellermann *et al.* *N. Engl. J. Med.* **329**, 1084–1091; 1993) and a 4.8-fold greater risk of suicide (A. L. Kellermann *et al.* *N. Engl. J. Med.* **327**, 467–472; 1992). Ever since, Congress has included in annual spending laws the stipulation that none of the CDC's injury-prevention

funds “may be used to advocate or promote gun control”.

The gun lobby's reach grew still wider this year, when the ban was extended to all agencies in the Department of Health and Human Services, including, most prominently, the NIH. The agency, to its credit, has chosen to read the ban narrowly. “The NIH supports research and public health education programs on injury prevention and violence reduction,” it said in a statement. “This effort includes programs related to firearm violence, which is a public health concern.”

Even so, the work that the NIH does support in this arena is limited. A search of the agency's grant database for the word ‘firearm’ returns just five projects, funded at a combined total of \$2.6 million in 2011. One of these looks at the relationship between acute alcohol use and different methods of suicide, including the use of firearms. Another aims to create a training and education resource for families of children with traumatic brain injury.

Like any sound public policy, rational decisions on firearms cannot be born in a scientific vacuum. There is a desperate need for peer-reviewed research to address even basic questions, such as whether there is a way to use the registration and licensing of gun owners to reduce the associated fatalities — which totalled 31,347 in the United States in 2009, the most recent year for which data are available. It is incumbent on scientists and the public to insist to their law-makers that research on such rudimentary questions is not sacrificed on the altar of politics. If the politicians do not hear this message forcefully and regularly, the chilling effect of special interests on research into charged but crucial questions in any number of policy areas will only grow. ■

Take a look

Enjoy Curiosity on Mars. We may not see its like again.

By now, many will have seen the image of Curiosity's descent taken by the Mars Reconnaissance Orbiter (MRO), which captured the rover one minute before its successful landing on Mars' Gale Crater earlier this week. One robot taking a picture of another robot. Above another planet.

Four years ago, the MRO snapped another lander, Phoenix, in a similar situation minutes before it touched down on the red planet. But the two missions are very different. Phoenix got to Mars on a modest US\$420-million budget and lasted only five months. Curiosity, at a cost of \$2.5 billion, has a rugged design and a nuclear power source that should mean its 400 scientists will be gainfully employed for years

to come (see page 137). The rovers' objectives are vastly different, too. Phoenix was stuck scraping for ice in one completely flat spot. Curiosity will climb its 5.5-kilometre-high target — the mountain Aeolis Mons, also known as Mount Sharp — and attempt to unpack the hundreds of millions of years of Martian geological history it contains.

There is something that binds Phoenix, Curiosity and the MRO together besides photography. Since NASA revamped its Mars programme in 2000, it has made a concerted effort to launch regular probes to a Solar System neighbour that is a mere nine-month rocket ride away. Four years before Phoenix, NASA landed the Spirit and Opportunity rovers there. And four years before the 2005 MRO, the agency had launched the Odyssey orbiter. Curiosity and its landing system represent the culmination of technological expertise and lines of scientific enquiry that have been nurtured for more than a decade.

Will the agency get the chance to use these powerful capabilities again? That depends largely on the whims of politicians and the economy, which are both conspiring to dim the future of NASA's Mars programme. It is a good thing that Curiosity could survive for a decade, because it is unlikely that the world will see anything like it for a while. ■

with words and phrases mined from the advert. A computer will almost certainly do the initial screen to weed out non-viable candidates and assign ratings; applicants shouldn't risk their submission being deleted just because it didn't contain the appropriate keywords. "Give them every reason to screen you in instead of screening you out," says Tringali, who adds that résumés and CVs for industry should not be overly technical.

In the interest of brevity, industrial applications should not include appendices, says Blackford. Terry Jones, senior careers consultant at the Careers Group, University of London, agrees. "Some applicants think that hiring managers will be happier with a much longer account. But people are busy," he stresses. "You need to get over some key points pretty quickly. It's about clarity, not about endless detail." Including irrelevant information about publications, grants, awards and presentations could also send the wrong message: "If industry sees someone with a huge publications appendix, they may think this person is still hanging on to academic culture," says Blackford.

Those applying to positions in Germany, Austria and parts of Switzerland need to be aware of cultural differences, says Barbara Janssens, PhD career manager for the German Cancer Research Center in Heidelberg. If an advert is in German, employers in those countries expect CVs in the same language — not in English. CVs should include a professional photograph of the applicant, she says, and must be signed and dated. They should also include personal details such as date of birth and marital status, and copies or PDFs of diplomas and certificates.

When applying for scientific positions in most other European nations, it is safe to send application materials in English, unless the advert is in another language. In that case, applicants should contact the employer to learn what language to use.

Ultimately, say careers advisers, applicants need to suss out potential employers' expectations for format, language and other uncertainties by reading the advert, checking with mentors, reaching out to contacts who work for the employer and asking the employer themselves. The most brilliant research accomplishments can't work in an applicant's favour if the CV or résumé goes unread, as Sharon Milgram, director of the Office of Intramural Training and Education at the US National Institutes of Health in Bethesda, Maryland, points out. "Don't blow your chances," she says, "by not giving me what I want." ■

Karen Kaplan is assistant Careers editor for Nature.

TURNING POINT

Brian Fisher

Brian Fisher, an entomologist and curator at the California Academy of Sciences in San Francisco, has maintained funding from the US National Science Foundation (NSF) since 1994 to collect and study ants from around the world. He has identified more than 1,000 species and studied their evolution. After a few US politicians suggested that his grants are wasteful government spending, he started considering different ways to fund his research. This year, Fisher found some success with crowd-funding through a website called Petridish.



How did your research become politicized?

It started in 2010, when Senator John McCain (Republican, Arizona) listed my field work in East Africa — collecting ants and sharing their photos and information on the AntWeb website — as number six of what he considered the top 100 most wasteful projects funded by the American Recovery and Reinvestment Act of 2009. Since then, the project has been cited in at least six different Republican campaign commercials as an example of how President Barack Obama's administration wastes money. It has been interesting to get raked over the coals. But what caught me most off guard was that the critics, such as conservative radio personalities, weren't necessarily focused on funding for ants; they were questioning whether the government should fund science at all.

Has there been any fallout for your career?

I have a big research project, with 25 people involved worldwide. I'm worried about the next NSF funding cycle, and the negative publicity doesn't help. So I've been looking at alternative sources of funding, including crowd-funding: small contributions from online donors.

How has science funding changed since you got your PhD?

Scientists have become more like entrepreneurs, having to seek many sources of support. Ant research has always been on the fringe, getting by on crumbs of funding. But you used to be able to sustain your career on NSF funding. Now scientists need a portfolio of options.

How did you learn to create such a portfolio?

I dropped out of my first graduate programme, in biology, because the only money available required me to work on a project that I wasn't passionate about. I spent a year incubating ideas and writing grant applications — figuring out how to raise my own money — so that I could pursue research on ant diversity. Since then, I have raised well over US\$750,000 from

unconventional sources, including private donors, corporations and foundations, to create the Madagascar Biodiversity Center in Antananarivo, which identifies land for conservation and catalogues local species. I've also been able to create Ant Course: a field course offered in different countries to teach students about ant taxonomy and field-research techniques.

Describe your experience on Petridish.

I wanted to secure at least \$10,000 in funding to visit a remote, rugged, pristine forest in northwestern Madagascar, to collect ant species before the habitat is converted for cattle raising. My project was posted online for 45 days and I landed 94 backers — ranging from one who pledged more than \$5,000 to 48 who pledged about \$20 each. This was my first experience with crowd-funding and it was really hard, especially shooting the requisite video pitch. I'm used to investing three weeks of blood and sweat writing an NSF grant application, but speaking directly to the public was very different.

How might crowd-funding help science?

Scientists need alternative sources of money now, but that is just one of the benefits. Crowd-sourcing helps to democratize science — the websites let amateur scientists participate. There is a public-relations aspect — you have to make clear the relevance of your research. I think every graduate student should try to get funding in this way, because the emphasis is on communication. They would need to focus their questions and make a pitch, but a few thousand dollars could be enough to support them. Graduate students need to learn how to advocate for their field — you can't just hide inside the ivory tower. The walls are gone. ■

INTERVIEW BY VIRGINIA GEWIN

CAREERS

TURNING POINT Political criticism prompts entomologist to seek crowd-funding p.243

CAREERS BLOG The latest discussions and scientific-careers issues go.nature.com/z8g4a7

NATUREJOBS For the latest career listings and advice www.naturejobs.com



JOB APPLICATIONS

Straight to the top of the pile

The distinctions between a lengthy, technical CV and a snappy résumé can make a big difference in a job search.

BY KAREN KAPLAN

Competition for research positions is fierce, and first impressions are crucial. In most cases, a written job application represents a potential employer's first exposure to an applicant. But assembling a list of qualifications and achievements is no simple matter. Should it be a short, snappy résumé or an exhaustive CV? How should it look? How technical should it be?

Early-career researchers need to know and

understand the expectations of employers in different countries and sectors. The content, format, design and language of the application materials can influence a candidate's chances of nabbing an interview. The wrong type of document, or one that has too much, too little or irrelevant information, can propel an application to the bottom of the pile — or disqualify it altogether. Given the state of the global economy, applicants can't afford to commit faux pas that could jeopardize their chances.

In all countries, the CV and résumé provide

an overview of an applicant's experience and achievements, and candidates need to know what to include, how to format the document and how to organize and present it visually. But accepted practices differ. In the United States and Canada, a CV is comprehensive, whereas a résumé is concise. In the United Kingdom and the European Union (EU), 'CV' and 'résumé' are generally synonymous, but there are variations. Researchers need to know when one, the other or a hybrid is most suitable.

NATIONAL DIFFERENCES

A US-style CV is usually an exhaustive, chronological timeline of the applicant's education, career and accomplishments. It starts with educational history and research experience, and includes a full record of awards, publications, grants received, conferences presented at, peer-review experience, committee work and other activities. Most early-career scientists have CVs longer than ten pages, and not all the content will be relevant to the position sought.

A US-style résumé, by contrast, is a succinct and tailored summary of the applicant's work and education, which focuses on his or her qualifications for the position. It should grab the reader's attention: according to a study published this year, US recruiters spend an average of just six seconds reading a résumé before deciding whether to pass it on or bin it (W. Evans, *Eye Tracking Online Metacognition: Cognitive Complexity and Recruiter Decision Making*; TheLadders, 2012). Presentation and structure are therefore paramount: a résumé's first page must highlight the applicant's relevant skills, knowledge and capabilities.

The same study found that recruiters spend most of those six seconds on the applicant's name, educational history, current position and most recent past position. The résumé should therefore keep descriptions of pertinent career and training experience and significant grants, publications and awards very succinct. Most résumés, even for mid-career scientists, should not exceed three pages. "The résumé is typically shorter and simpler than a CV because a non-scientist is often the first reader," says Joe Tringali, managing director of Tringali and Associates, a recruitment consultancy in Manchester, New Hampshire. "It's slightly salesy — it's a marketing document, with less emphasis on publications and more on techniques and skills."

It is important for international researchers seeking US positions to note that résumés ►

K. REINICK/SHUTTERSTOCK

► should not include personal information or a personal photograph.

Appearance and layout are crucial for US résumés, although they are relatively insignificant for CVs in comparison with content. In a résumé, the applicant's name and contact information should be in large type at the top of the first page. There should be wide margins at the top, bottom and sides of each page and separating each entry; the main text must be no smaller than 11 point; and the font should be clean and easy to read.

THE RIGHT TOOLS FOR THE JOB

In general, say US hiring managers and careers advisers, applicants should send a full-length CV for research and grant applications (including fellowships) in academia, government and non-profit organizations, whereas a résumé is best for non-research posts in government, the non-profit sector and industry. For academic research posts that involve some teaching, such as at small liberal-arts colleges, or for academic or industrial postdoctoral research, the best option is a hybrid style that includes relevant skills, capabilities and knowledge on the first page. It should also contain information such as full or partial listings of grants, awards, publications and experience as a reviewer or on committees, and could total between three and eight pages. If there is a great deal more information on publications, posters, conference presentations, books authored and so on that is relevant to the position being sought, applicants can send the hybrid but append a note saying that a full CV is available.

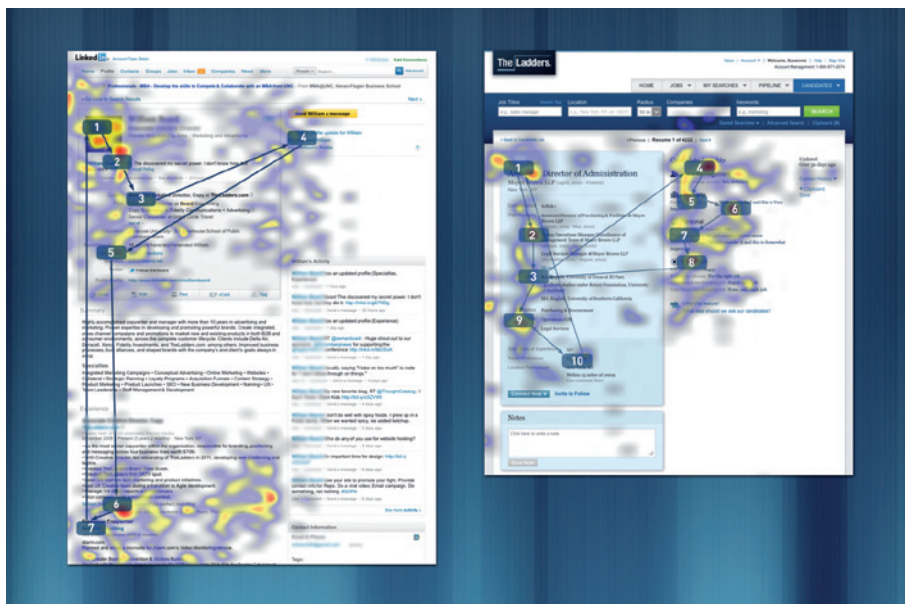
The hybrid can also work well for professional-networking websites such as LinkedIn, where work and educational history should be informative and career experience clear, but an exhaustive list of publications, presentations, committee work and other experience is unnecessary. It is important, however, to include relevant keywords online, because many recruiters search for job candidates on networking sites and use keywords to refine and narrow their search. Careers advisers say that here, too, it is often best to link to the full CV, or add a note saying that it is available on request.

For US industrial-research applications, a résumé is almost obligatory — unless the applicant knows for certain that they are sending their submission directly to a scientist. If an advert requests a CV, résumé or both, it is best



Applicants need to be aware of cultural differences in different countries.

Barbara Janssens



Studies show that recruiters focus on CV and résumé hotspots (red) such as name and most recent job.

to send a résumé, but indicate that a full CV is available. If the advert provides no instruction and gives a generic e-mail address, such as careers@xyzbio.com, a non-scientist screener will usually view the document first, and will expect a résumé.

In the United Kingdom and the EU, the long-form CV is rare and employers in all sectors are used to a style more or less identical to the US hybrid version. Applicants to UK positions should lead with contact information and personal data such as nationality, visa status, age and whether they have a driving licence, says Sarah Blackford, head of education and public affairs at the Society for Experimental Biology in London. She notes, however, that age is optional, and there should never be a photograph. Academic CVs should include references to attached appendices in which the applicant can more fully explain their research experience, publications, grants and conference presentations.

The European Commission is trying to standardize job-application formats using an online tool for uploading CVs (go.nature.com/yxm9r1). Unlike in the United Kingdom, CVs in the EU should include a photograph, says Emilia Daniłowicz-Luebert, an immunology PhD student at the Humboldt University in Berlin and team leader for the Careers in Life Sciences Project of the Young European Bio-tech Network in Bertinoro, Italy.

In China, CV formats are similar to the long-form US style, says Zhou Zhonghe, director of the Institute of Vertebrate Paleontology and Paleoanthropology in Beijing. He says that most employers offer an online tool for uploading.

Applicants can ensure that they are prepared for any eventuality by keeping multiple versions of résumés and CVs. Matt Hepworth, a UK-born immunologist at the Humboldt

University who is seeking a second postdoc position, has two versions of a hybrid, which he has sent out for industrial and academic jobs in the United States, the EU and the United Kingdom.

The first page of both versions includes a summary of his research focus and its implications for disease, but the industrial version also has a section highlighting his relevant skills and the techniques and protocols that he has mastered. The academic version lists conference presentations, travel grants and journals for which he is a peer reviewer. Hepworth's academic version is five pages long; the industrial version is three-and-a-half. "I really tried to restrict the length, even on the academic one," says Hepworth. Staying concise seems to have paid off — Hepworth is negotiating for a US postdoc job that he hopes will eventually become a tenure-track research post.

TAILOR TO FIT

Applications should be tailored for the post of interest. When the US long-form CV is called for, candidates should tailor through their cover letter; otherwise, the CV itself should be adjusted. UK industrial-research applicants should consider the description of the position in the advert when composing their research summaries, highlights and all text outlining current and past research positions and fellowships. "Use your positions as headlines and then demonstrate through your descriptions that you can meet the needs of that business," says Alison Mitchell, deputy director of Vitae, a research-career-support organization in Cambridge, UK.

In all countries, it is likely that the first viewer of an industrial application will not be a scientist, so an applicant's résumé (in the United States) or CV (in the United Kingdom and EU) should be peppered liberally

with words and phrases mined from the advert. A computer will almost certainly do the initial screen to weed out non-viable candidates and assign ratings; applicants shouldn't risk their submission being deleted just because it didn't contain the appropriate keywords. "Give them every reason to screen you in instead of screening you out," says Tringali, who adds that résumés and CVs for industry should not be overly technical.

In the interest of brevity, industrial applications should not include appendices, says Blackford. Terry Jones, senior careers consultant at the Careers Group, University of London, agrees. "Some applicants think that hiring managers will be happier with a much longer account. But people are busy," he stresses. "You need to get over some key points pretty quickly. It's about clarity, not about endless detail." Including irrelevant information about publications, grants, awards and presentations could also send the wrong message: "If industry sees someone with a huge publications appendix, they may think this person is still hanging on to academic culture," says Blackford.

Those applying to positions in Germany, Austria and parts of Switzerland need to be aware of cultural differences, says Barbara Janssens, PhD career manager for the German Cancer Research Center in Heidelberg. If an advert is in German, employers in those countries expect CVs in the same language — not in English. CVs should include a professional photograph of the applicant, she says, and must be signed and dated. They should also include personal details such as date of birth and marital status, and copies or PDFs of diplomas and certificates.

When applying for scientific positions in most other European nations, it is safe to send application materials in English, unless the advert is in another language. In that case, applicants should contact the employer to learn what language to use.

Ultimately, say careers advisers, applicants need to suss out potential employers' expectations for format, language and other uncertainties by reading the advert, checking with mentors, reaching out to contacts who work for the employer and asking the employer themselves. The most brilliant research accomplishments can't work in an applicant's favour if the CV or résumé goes unread, as Sharon Milgram, director of the Office of Intramural Training and Education at the US National Institutes of Health in Bethesda, Maryland, points out. "Don't blow your chances," she says, "by not giving me what I want." ■

Karen Kaplan is assistant Careers editor for Nature.

TURNING POINT

Brian Fisher

Brian Fisher, an entomologist and curator at the California Academy of Sciences in San Francisco, has maintained funding from the US National Science Foundation (NSF) since 1994 to collect and study ants from around the world. He has identified more than 1,000 species and studied their evolution. After a few US politicians suggested that his grants are wasteful government spending, he started considering different ways to fund his research. This year, Fisher found some success with crowd-funding through a website called Petridish.



How did your research become politicized?

It started in 2010, when Senator John McCain (Republican, Arizona) listed my field work in East Africa — collecting ants and sharing their photos and information on the AntWeb website — as number six of what he considered the top 100 most wasteful projects funded by the American Recovery and Reinvestment Act of 2009. Since then, the project has been cited in at least six different Republican campaign commercials as an example of how President Barack Obama's administration wastes money. It has been interesting to get raked over the coals. But what caught me most off guard was that the critics, such as conservative radio personalities, weren't necessarily focused on funding for ants; they were questioning whether the government should fund science at all.

Has there been any fallout for your career?

I have a big research project, with 25 people involved worldwide. I'm worried about the next NSF funding cycle, and the negative publicity doesn't help. So I've been looking at alternative sources of funding, including crowd-funding: small contributions from online donors.

How has science funding changed since you got your PhD?

Scientists have become more like entrepreneurs, having to seek many sources of support. Ant research has always been on the fringe, getting by on crumbs of funding. But you used to be able to sustain your career on NSF funding. Now scientists need a portfolio of options.

How did you learn to create such a portfolio?

I dropped out of my first graduate programme, in biology, because the only money available required me to work on a project that I wasn't passionate about. I spent a year incubating ideas and writing grant applications — figuring out how to raise my own money — so that I could pursue research on ant diversity. Since then, I have raised well over US\$750,000 from

unconventional sources, including private donors, corporations and foundations, to create the Madagascar Biodiversity Center in Antananarivo, which identifies land for conservation and catalogues local species. I've also been able to create Ant Course: a field course offered in different countries to teach students about ant taxonomy and field-research techniques.

Describe your experience on Petridish.

I wanted to secure at least \$10,000 in funding to visit a remote, rugged, pristine forest in northwestern Madagascar, to collect ant species before the habitat is converted for cattle raising. My project was posted online for 45 days and I landed 94 backers — ranging from one who pledged more than \$5,000 to 48 who pledged about \$20 each. This was my first experience with crowd-funding and it was really hard, especially shooting the requisite video pitch. I'm used to investing three weeks of blood and sweat writing an NSF grant application, but speaking directly to the public was very different.

How might crowd-funding help science?

Scientists need alternative sources of money now, but that is just one of the benefits. Crowd-sourcing helps to democratize science — the websites let amateur scientists participate. There is a public-relations aspect — you have to make clear the relevance of your research. I think every graduate student should try to get funding in this way, because the emphasis is on communication. They would need to focus their questions and make a pitch, but a few thousand dollars could be enough to support them. Graduate students need to learn how to advocate for their field — you can't just hide inside the ivory tower. The walls are gone. ■

INTERVIEW BY VIRGINIA GEWIN

THIS WEEK

EDITORIALS

WORLD VIEW Arab revolutions positive for science — for now **p.131**

FOOD OF LOVE Frisky flies get picked off by hungry bats **p.132**

HUMANITIES Language and art around for longer than we thought **p.133**



Who calls the shots?

US law-makers need to encourage research on firearms-related violence so that gun laws can be based on facts rather than ideology.

The shooting spree in a Colorado cinema last month that killed 12 people and injured 58 has provoked many questions and much soul-searching. Some reports have even suggested that because the perpetrator, James Holmes, was until just weeks earlier a graduate student supported by a US National Institutes of Health (NIH) training grant, the biomedical agency is somehow implicated.

The insinuation is ludicrous, but the attention it received speaks volumes about the political reluctance in the United States to address the laws that made it possible for Holmes to obtain his arsenal of firearms. In this climate, discussions of the multiple murders sounded all too often like descriptions of the random and inevitable carnage caused by a tornado or an earthquake.

Natural disasters, truly unavoidable events, can be combated with science. The US Geological Survey, for instance, has some 250 employees dedicated to the assessment of earthquake hazards.

There is no such US government effort for research on firearms — the National Rifle Association has helped see to that. The lobby group that represents gun owners began to squash scientific efforts in 1996, when, using proxies in Congress, it shut down a fledgling, US\$2.6-million gun-violence research effort by the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia.

Among other things, that work revealed that people living in homes where there was a gun faced a 2.7-fold greater risk of homicide (A. L. Kellermann *et al.* *N. Engl. J. Med.* **329**, 1084–1091; 1993) and a 4.8-fold greater risk of suicide (A. L. Kellermann *et al.* *N. Engl. J. Med.* **327**, 467–472; 1992). Ever since, Congress has included in annual spending laws the stipulation that none of the CDC's injury-prevention

funds “may be used to advocate or promote gun control”.

The gun lobby's reach grew still wider this year, when the ban was extended to all agencies in the Department of Health and Human Services, including, most prominently, the NIH. The agency, to its credit, has chosen to read the ban narrowly. “The NIH supports research and public health education programs on injury prevention and violence reduction,” it said in a statement. “This effort includes programs related to firearm violence, which is a public health concern.”

Even so, the work that the NIH does support in this arena is limited. A search of the agency's grant database for the word ‘firearm’ returns just five projects, funded at a combined total of \$2.6 million in 2011. One of these looks at the relationship between acute alcohol use and different methods of suicide, including the use of firearms. Another aims to create a training and education resource for families of children with traumatic brain injury.

Like any sound public policy, rational decisions on firearms cannot be born in a scientific vacuum. There is a desperate need for peer-reviewed research to address even basic questions, such as whether there is a way to use the registration and licensing of gun owners to reduce the associated fatalities — which totalled 31,347 in the United States in 2009, the most recent year for which data are available. It is incumbent on scientists and the public to insist to their law-makers that research on such rudimentary questions is not sacrificed on the altar of politics. If the politicians do not hear this message forcefully and regularly, the chilling effect of special interests on research into charged but crucial questions in any number of policy areas will only grow. ■

Take a look

Enjoy Curiosity on Mars. We may not see its like again.

By now, many will have seen the image of Curiosity's descent taken by the Mars Reconnaissance Orbiter (MRO), which captured the rover one minute before its successful landing on Mars' Gale Crater earlier this week. One robot taking a picture of another robot. Above another planet.

Four years ago, the MRO snapped another lander, Phoenix, in a similar situation minutes before it touched down on the red planet. But the two missions are very different. Phoenix got to Mars on a modest US\$420-million budget and lasted only five months. Curiosity, at a cost of \$2.5 billion, has a rugged design and a nuclear power source that should mean its 400 scientists will be gainfully employed for years

to come (see page 137). The rovers' objectives are vastly different, too. Phoenix was stuck scraping for ice in one completely flat spot. Curiosity will climb its 5.5-kilometre-high target — the mountain Aeolis Mons, also known as Mount Sharp — and attempt to unpack the hundreds of millions of years of Martian geological history it contains.

There is something that binds Phoenix, Curiosity and the MRO together besides photography. Since NASA revamped its Mars programme in 2000, it has made a concerted effort to launch regular probes to a Solar System neighbour that is a mere nine-month rocket ride away. Four years before Phoenix, NASA landed the Spirit and Opportunity rovers there. And four years before the 2005 MRO, the agency had launched the Odyssey orbiter. Curiosity and its landing system represent the culmination of technological expertise and lines of scientific enquiry that have been nurtured for more than a decade.

Will the agency get the chance to use these powerful capabilities again? That depends largely on the whims of politicians and the economy, which are both conspiring to dim the future of NASA's Mars programme. It is a good thing that Curiosity could survive for a decade, because it is unlikely that the world will see anything like it for a while. ■



Arab liberals must stay in the game

Islamist academics are gaining power in the Middle East and North Africa. But to build science needs liberal input, argues Ehsan Masood.

Back in 2006, I wrote a piece for *Nature* on what Islamist science policy might look like. It was for a special issue on Islam and science, which carried the cover-line: 'Must the Muslim world stay science poor?'

Would the environment for science conceivably improve, the issue asked, if more countries elected Islamist physicians and scientists to parliaments and presidencies? Islamists tend to be economic liberals but social ultra-conservatives, and are often dogmatic in any clash between belief and reason. What, then, would Islamist government mean for free thinking and for international scientific cooperation?

In recent weeks, Twitter users and others following or involved in the wave of Arab revolutions in the past two years have rediscovered my article (E. Masood *Nature* 444, 22–25; 2006) and have questioned me about the impact of the political changes. They have asked, for example, to what extent Islamist political parties taking office will become authoritarian and lose tolerance for dissent.

Now is a good time to revisit the issue and to offer a brief update. Governments affiliated to the pan-Arab Islamist organization the Muslim Brotherhood are in office in Tunisia and Morocco, and in June took the biggest prize of all: Egypt.

Five years ago, Kemal Helbawy, once an organizer of Brotherhood exiles in Europe (the group has been banned in Egypt for most of its existence) told me that Islamist science policy would have three goals: supporting national defence and security, ensuring quality of life and 'proving' the miraculous nature of the Islamic faith. Centuries of science, distilled into three bullet points. But how much are Islamists being forced to change that approach, now that they have swapped protest for power? How do the predictions of 2006 look now? And what does that mean for current predictions about science in these nations?

In 2006, the special issue of *Nature* forecast that more Islamist governments were likely to take office after free elections, something that few Western policy-makers were willing to countenance. Even if they knew it, they weren't saying so in public. Where the article writers were wrong was in thinking that Islamist governments would close down and prevent scientific cooperation. Thankfully, there is little evidence of this so far — if anything, international scientific contacts may increase. Organizations including the British Council and the European Commission, which already fund research cooperation in the Middle East and North Africa, are responding to questions from Libya, Morocco and Tunisia on how to modernize research and higher education.

The future of US scientific links to Islamic nations — tentatively strengthened in 2009 with the appointment of science envoys — could rest

on the foreign policy of the winner of this year's US presidential election, and on how the post-revolution Islamist governments evolve.

Egypt, for one, seems to be following Helbawy's suggestions for science policy. President Mohammed Morsi, a US-trained former professor of engineering at Zagazig University in Egypt, is certainly addressing the 'quality of life' goal: he has asked his science ministry to revive research into applied urban topics such as poor sanitation, atmospheric pollution and traffic congestion.

Another thing that *Nature* got wrong in 2006 was an assumption that Islamist governments would have little appetite or tolerance for freedoms. The articles assumed that there would be little or no freedom to think, to speak and to ask questions of those in power, which are crucial ingredients for successful science.

This does not seem to be the case for now, from what I have seen on trips to Egypt and Morocco.

In April, I attended a seminar in Alexandria on the future of Egyptian science policy. The speakers included a junior minister for science from Cairo. But for the 200-strong audience of scientists and students, mostly under 30 years old, his status (and the power it brought) did not intimidate.

The room fizzed with anger. Speaker after speaker grabbed the microphone to bark at the minister: "How come you still have a job?"; "What do you earn?"; "I want to do a PhD. What will you do to improve my job prospects?" A few years ago, such an event might have resulted in arrests and torture, but now the tables have turned. The man from the ministry had to behave as a servant of the people and try to answer. He wasn't an Islam-

ist politician, but his elected bosses are, and it seems that they (so far) are willing to let Egypt's youth ask any question of anyone in authority.

It may not last. Religious movements by their very nature tend to be authoritarian, and once the Islamists have no need for the first generation of revolutionaries, they may revert to type.

To keep this from happening, or to delay it, the academics and other political activists who gave birth to revolutions will need to persevere with the political game. There is a risk that many of them, disillusioned with the slow pace of change, will return to their day jobs and leave the political field clear for the much better-organized Islamists.

If the Islamists are genuinely committed to using the best available knowledge to build their nations, they can learn plenty from what has gone wrong in Iran, Pakistan, Saudi Arabia, Turkey and elsewhere. And perhaps most important of all, they must see value in creating relationships with those citizens who do not sit in the rows of the pious. ■

Ehsan Masood is editor of *Research Fortnight* and author of *Science and Islam: A History*. He is based in London.
e-mail: news@researchresearch.com

THE ACADEMICS WHO
GAVE BIRTH TO
REVOLUTIONS
WILL NEED TO
PERSEVERE WITH THE
POLITICAL
GAME.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/krkod9

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

ASTRONOMY

Star dines on young planet

Astronomers report the discovery of a possible extrasolar planet just a few million years old — which poses problems for some models of planetary formation.

The planet — the youngest yet seen to cross the face of its host star — is about 5.5 times the mass of Jupiter and takes just 11 hours to orbit its 2.7-million-year-old star. The planet is so close to the star that Julian van Eyken of the California Institute of Technology in Pasadena and his team suggest that it may be in the process of being consumed. Moreover, the young age of this extrasolar system challenges astronomical models that require up to ten million years for planetary formation to occur.

Astrophys. J. 755, 42 (2012)

GEOLOGY

Plants changed water cycle

A doubling of carbon dioxide levels some 200 million years ago may have reduced plants' uptake and release of water — drastically altering local water cycles and leading to a decrease in animal biodiversity.

Margret Steinthorsdottir, now at Stockholm University, and her team examined 91 fossil plants from eastern Greenland, spanning the transition between the Triassic

and Jurassic periods. The researchers measured the fossils' stomata — tiny holes through which plants vent water (**stoma pictured**) — and found that their density and size decreased over the Triassic–Jurassic transition. This suggests that the volume of water released by plants in a process called transpiration fell by 50–60% during a time marked by mass species extinctions and high levels of atmospheric carbon dioxide.

Sediment analysis revealed that the drop in transpiration coincided with increased water run-off and erosion, suggesting that the change may have reduced soil quality, contributing to a decline in biodiversity.

Geology <http://dx.doi.org/10.1130/G33334.1> (2012)

CLIMATE SCIENCE

Abrupt changes in Greenland ice cycles

An analysis of aerial photographs suggests that ice loss in northwestern Greenland results from discrete events driven by changes in ocean and atmospheric temperature.

Kurt Kjær at the University of Copenhagen's Natural History Museum and his team used photographs dating back to 1985 to create a three-dimensional model of ice loss and gain.

Rather than exhibiting a uniform melt rate, the model revealed that ice loss peaked in two periods: 1985–93 and 2005–10. The researchers linked this ice loss to warmer oceans and higher summer air temperatures. Ice-sheet models must account for such variability if they are to produce reliable forecasts.

Science 337, 569–573 (2012)

CHEMISTRY

Nanorods all in a row

Tiny carpets of gold rods, all standing upright, can be manufactured more easily and accurately thanks to a method exploiting capillary action. These nanorods have precise optical properties that can be used in sensors and solar energy harvesting.

Udo Bach at Monash University in Clayton, Australia, and his team patterned silica-coated wafers with gold squares 2 micrometres wide and 4 micrometres apart. The wafers are placed in tubes with an aqueous solution of gold nanorods modified with thiol and polyethyleneglycol (PEG).

When the solvent evaporates, the nanorods crystallize upwards from the gold squares. Thiol and PEG molecules bind the rods together as capillary action from the evaporating solvent drags them upright. The process can be controlled easily by changing the solution concentration and evaporation temperature.

Angew. Chem. Int. Edn. <http://dx.doi.org/10.1002/anie.201204609> (2012)

ZOOLOGY

Bats sound out frisky flies

Hungry bats can tune in to the sound of flies mating to pick out tiny prey that they would otherwise be unable to detect.

Stefan Greif at the Max



J. BALOG/AURORA/GETTY

Planck Institute for Ornithology in Seewiesen, Germany, and his team analysed videos of Natterer's bats (*Myotis nattereri*; pictured) feeding on *Musca domestica* flies on a cowshed ceiling. Thousands of lone flies walked across the area without being attacked. However, copulating pairs were attacked 5.3% of the time. Bats also attacked loudspeakers playing the distinctive ultrasound buzzing produced by copulating flies.

This is the first identified mechanism that supports theories that copulation can leave animals more vulnerable to attack.

Curr. Biol. 22, R563–R564 (2012)

ARCHAEOLOGY

Modern thinking gets older

Modern human behaviour underlying cultural innovations such as language and art might have begun in southern Africa thousands of years earlier than assumed.

Evidence for symbolic behaviour, such as shell beads, appeared at least 80,000 years ago in southern Africa. This behaviour then seemingly disappeared and did not return until roughly 20,000 years ago — when humans with cultural links to modern San hunter-gatherers began to produce engraved bones and other complex artefacts.

However, Francesco d'Errico at the University of Bordeaux in France, Paola Villa at the University of Colorado in Boulder and their teams suggest that antecedents to San culture in fact appeared much earlier. Radiocarbon dating of seashell and ostrich eggshell beads (pictured), complex resins and an ochre-stained point previously excavated from Border Cave in South

Africa suggests that the artefacts are up to 44,000 years old. Around this time, humans living near Border Cave also began to produce double-faced stone blades and flint arrow points — consistent with the emergence of modern symbolic behaviour, the authors say.

Proc. Natl Acad. Sci. USA <http://dx.doi.org/10.1073/pnas.1204213109>; <http://dx.doi.org/10.1073/pnas.1202629109> (2012)

MEDICINE

Neighbours join the resistance

Normal cells near cancerous ones can support their malignant neighbours by secreting proteins in response to anticancer drugs.

Peter Nelson at the Fred Hutchinson Cancer Research Center in Seattle, Washington, and his team found that chemotherapy triggered benign cells near prostate tumours to secrete a signalling protein called WNT16B. This promoted tumour growth and could help tumours to become resistant to therapy. Targeting the regulators of WNT16B, or other components of the tumour microenvironment, could be routes to developing anti-resistance drugs.

Nature Med. <http://dx.doi.org/10.1038/nm.2890> (2012)

BIOLOGY

Pregnancy alters gut microbes

Pregnancy triggers a radical shift in human gut microflora, moving the body towards a diabetes-like condition.

Ruth Ley at Cornell University in Ithaca, New York, and her team analysed the faecal bacteria of 91 pregnant women. During the first trimester, these women's bacterial populations were

COMMUNITY CHOICE

The most viewed papers in science

NEUROBIOLOGY

Diabetes drug boosts neuron growth

HIGHLY READ
on www.cell.com/cell-stem-cell in July

Growing evidence suggests that the mammalian brain recruits adult neural stem cells in an attempt to repair diseased or injured neurons. Drugs that can spur on this recruitment are highly sought after.

One candidate is the widely used diabetes drug metformin, which activates an enzyme called aPKC. In the brain, this protein's action on another protein, CBP, is essential for optimal specialization of neural precursor cells.

Freda Miller at the Hospital for Sick Children in Toronto, Canada, and her team showed that metformin also activated the aPKC–CBP pathway in cultured mouse and human neural precursors, promoting neuron generation. Moreover, it enhanced the generation of new neurons in the brains of live adult mice. Crucially, the change seemed to confer a benefit: adult mice treated with metformin showed better spatial-memory formation in a water maze compared with controls.

Cell Stem Cell 11, 23–35 (2012)

similar to each other and to those of non-pregnant women. But by the third trimester, there were marked differences between the mothers-to-be, and their average gut make-up resembled that seen in a mouse model of metabolic syndrome — a collection of symptoms linked to diabetes.

When the microbiota from third-trimester women were transferred into mice, the animals gained more fat and became less sensitive to insulin than did mice that received first-trimester microbes. Mammals can manipulate their gut biology to trigger changes beneficial to a developing fetus, the researchers suggest.

Cell 150, 470–480 (2012)

For a longer story on this research, see go.nature.com/4v7cht

MATERIALS

SLIPS blitz biofilms

Coating solids with immobilized liquids could help to deal with dangerous bacterial biofilms.

Many bacterial species form durable slimes that can cover everything from

medical devices to ship hulls. Joanna Aizenberg at Harvard University in Cambridge, Massachusetts, and her team showed that slippery liquid-infused porous surfaces (SLIPS) — in which a liquid is 'locked' onto a solid through affinities between the two — can resist biofilm growth.

Whereas conventional slippery surfaces such as Teflon were colonized by biofilms within hours, SLIPS reduced the attachment of *Pseudomonas aeruginosa* biofilms by 99.6% over seven days. Similar success rates were seen against *Staphylococcus aureus* and *Escherichia coli*. SLIPS are stable under a variety of conditions and their anti-biofilm properties are not due to toxicity of the immobilized liquid. This makes them candidates for many anti-biofilm applications, say the authors.

Proc. Natl Acad. Sci. USA <http://dx.doi.org/10.1073/pnas.1201973109> (2012)

➔ **NATURE.COM**

For the latest research published by Nature visit:

www.nature.com/latestresearch



SEVEN DAYS

The news in brief

POLICY

Deforestation down

Deforestation in the Brazilian Amazon fell by more than 20% between August 2011 and July 2012, according to an early — and uncertain — analysis released on 2 August by Brazil's National Institute for Space Research. The preliminary figures are based on coarse satellite data, but suggest that deforestation is likely to hit a record low for the fourth season in a row. See go.nature.com/gzuthw for more.

H5N1 moratorium

Researchers should continue a self-imposed moratorium on lab studies that give new properties to the highly pathogenic avian influenza virus H5N1, according to Anthony Fauci, director of the US National Institute of Allergy and Infectious Diseases. Fauci was speaking at a meeting of flu researchers in New York City. The moratorium has been in place since late January; it was originally planned to last 60 days. See go.nature.com/3zwwq5 for more.

India's Mars hopes

The Indian cabinet has approved a small, 4.5-billion-rupee (US\$81-million) orbiting mission to Mars, which could be launched as early as November 2013. Approval came at a meeting on 3 August, according to the Indian national press agency. The orbiter would be the country's first interplanetary probe and would be carried by the same home-grown rocket used to launch its 2008 Moon probe, Chandrayaan-1.

Primate transport

Air China said on 31 July that it would stop shipments of non-human primates for research. The move followed sustained campaigning from

animal-rights group People for the Ethical Treatment of Animals (PETA), based in Norfolk, Virginia, which has led to many major air carriers refusing to fly primates bound for research centres (see *Nature* 483, 381–382; 2012). PETA says that China Eastern is the only major airline now known to be flying primates out of China — the country that last year transported more than 70% of the primates bound for US labs. See go.nature.com/ckhq93 for more.

BUSINESS

Gene-test regulation

Personal-genetics company 23andMe announced on 30 July that it was seeking

approval from the US Food and Drug Administration (FDA) for its genetic tests related to health — but not for those for non-medical information such as ancestry, or traits such as eye or hair colour. The firm, based in Mountain View, California, had previously maintained that it did not need FDA oversight because it provides information, not a medical service. See go.nature.com/cinowb for more.

GM patent win

One of the largest ever US patent settlements saw biotechnology giant Monsanto awarded \$1 billion on 1 August in a dispute with chemical company DuPont over

genetically engineered crops. Monsanto, based in St Louis, Missouri, argued that DuPont's agriculture subsidiary Pioneer Hi-Bred (now DuPont Pioneer) in Johnston, Iowa, infringed a Monsanto patent on Roundup Ready crops, which are resistant to the herbicide glyphosate. DuPont violated the patent by making soya beans that mingled the Monsanto trait with its own herbicide-resistant technology, the jury concluded. DuPont says that it will appeal the verdict.

Cancer vaccine

The biotech firm behind the first approved cancer vaccine will slash 41% of its workforce, after tepid sales of



A. DI MEO/EPA/CORBIS

Italian dog-breeding facility at risk

One of the largest suppliers of beagles (pictured) for mandatory drug testing in Europe could struggle to survive after an Italian court ordered its temporary closure and granted guardianship of the dogs to the animal-rights groups that filed charges of maltreatment. At an appeal hearing on 3 August, the Green Hill facility in Montichiari,

Italy, was told that its staff could re-enter the buildings, but that guardianship would not be returned. By 6 August, some 1,400 beagles had been placed in private homes; they will not be allowed to return to Green Hill because they could bring pathogens into the facility. See go.nature.com/t7am4p for more.

D. PARKER/SPL

its Provenge (sipuleucel-T) therapy for some prostate cancers. Dendreon, based in Washington, Seattle, saw Provenge approved in 2010 but doctors in the United States have not raced to prescribe the expensive (US\$93,000) treatment (see *Nature* 476, 376–377; 2011). Dendreon hopes to market the vaccine in Europe by 2013.

Alzheimer's setback

Research has been halted on a keenly watched experimental drug aimed at treating Alzheimer's disease after it failed two late-stage clinical trials. Pharmaceutical companies Johnson & Johnson, based in New Brunswick, New Jersey, and Pfizer, based in Groton, Connecticut, said on 6 August that they would no longer work on the monoclonal antibody bapineuzumab. It binds a peptide, amyloid- β , that may cause neurodegeneration in patients with Alzheimer's. But in two phase III trials — one reported on 23 July and one on 6 August — the drug did not prevent cognitive decline.

PEOPLE

Bernard Lovell dies

Physicist and radio astronomer Bernard Lovell, who founded the Jodrell Bank Observatory at the University of Manchester, UK, died on 6 August aged 98.



Lovell (pictured) directed the observatory from 1945 to 1980, and in 1957 oversaw the construction of its iconic telescope — then the world's largest fully steerable radio telescope — which opened in time to track the launch of the first artificial satellite, Sputnik 1. Lovell also worked on radar and cosmic rays; he was knighted in 1961 for his contributions to radio astronomy.

Fermilab change

Pier Oddone, the director of the Fermi National Accelerator Laboratory in Batavia, Illinois, announced on 2 August that he would step down from his post in July 2013. Since he started as the high-energy physics lab's fifth director in 2005, Oddone has overseen the final years of the lab's Tevatron particle collider and a shift in focus

to experiments involving neutrinos and high-intensity (rather than high-energy) collisions. See go.nature.com/2dn9hu for more.

Cold fusion death

Electrochemist Martin Fleischmann, who claimed to have discovered cold fusion, died on 3 August aged 85. He and Stanley Pons, both at the University of Utah in Salt Lake City, sparked worldwide controversy in 1989 when they announced to the world's media that they had fused deuterium atoms at room temperatures. Other scientists quickly debunked the experiments, and the claims have not been substantiated.

RESEARCH

Mars landing

NASA announced on 6 August that its Mars rover, Curiosity, had successfully landed in Gale Crater after an 8-month journey and a violent 7-minute fall through the planet's thin atmosphere. See page 137 and nature.com/curiosity for more.

Higgs papers

Researchers at ATLAS and the CMS, the two main physics experiments at the Large Hadron Collider near Geneva, Switzerland, posted their papers describing a new Higgs-boson-like particle to the online preprint server

arXiv.org on 31 July (ATLAS Collaboration <http://arxiv.org/abs/1207.7214> (2012) and CMS Collaboration <http://arxiv.org/abs/1207.7235>; 2012). The signals reported at the beginning of July are still there, and have gained significance well above five standard deviations. See go.nature.com/ivkhou for more.

Olympic biology

The UK government plans to transform the London Olympics drug-testing laboratory in Harlow into a national bioanalytics centre, Prime Minister David Cameron announced on 1 August. The MRC–NIHR Phenome Centre will use equipment purchased for anti-doping tests to probe biological samples for biochemicals such as proteins and metabolites, with the aim of analysing 100,000 samples per year from massive epidemiology projects. The Medical Research Council (MRC) and the National Institute of Health Research (NIHR) are each putting £5 million (US\$8 million) into the project.

Texas grant review

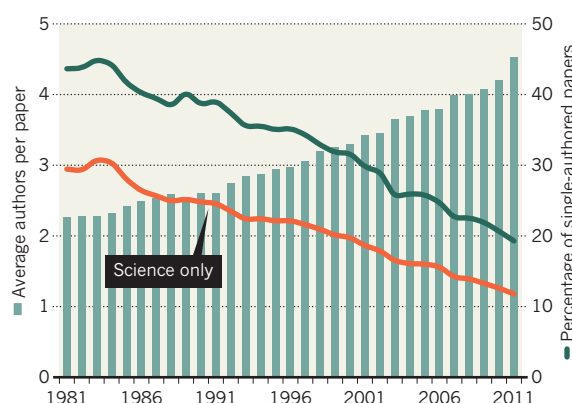
The Cancer Prevention and Research Institute of Texas (CPRIT) in Austin has appointed a compliance officer to review and monitor its grant-applications process, it announced on 2 August. Meeting for the first time since controversy erupted over its grant funding, the state-funded institute approved 45 grants, together worth more than US\$114 million. Seven of those grants were for multi-investigator projects that had been shelved back in March, when the CPRIT pushed through a \$20-million commercial 'incubator' award without scientific review. The unusual award prompted the resignation of the CPRIT's chief scientist and will now be re-reviewed. See go.nature.com/bzrzkc for more.

TREND WATCH

Research papers with a single author are becoming a rare breed, show data for 2011 from Thomson Reuters' *ScienceWatch* newsletter. Last year, less than 20% of papers in the sciences, social sciences, arts and humanities had single authors (see chart). In the sciences alone, single-author papers were down to 12%. At the other end of the spectrum, 2011 saw some 600 papers written by more than 100 authors, and 146 papers that had more than 1,000 authors — most of them in physics.

DECLINE OF THE SINGLE AUTHOR

The average number of authors on a research paper rose to 4.5 last year; fewer than one in five were single-authored.



SOURCE: THOMSON REUTERS/SCIENCEWATCH

NATURE.COM

For daily news updates see:
www.nature.com/news

NEWS IN FOCUS

BIOMEDICINE Japan bids to turn induced stem cells into therapies **p.139**

TECHNOLOGY Can graphene power a new Industrial Revolution in Britain? **p.140**

CLIMATE Heatwaves stoke debate about climate-weather link **p.143**



ENGINEERING 'Extreme mechanics' makes a virtue of failure **p.146**

NASA/JPL-CALTECH



An early image from Curiosity showed that the rover had landed facing the north flank of Mount Sharp.

PLANETARY SCIENCE

Mars rover sizes up the field

After a picture-perfect landing, Curiosity's science team ponders its first moves at Gale Crater.

BY ERIC HAND IN PASADENA

The mountain rises in the late-afternoon sun: a daunting challenge to the vehicle before it. In a photograph taken minutes after Curiosity landed on the surface of Mars, the rover's shadow already seems to be reaching for the distant slope that it was built to climb.

On 6 August, Curiosity — the rover for NASA's Mars Science Laboratory mission — arrived at its destination at the bottom of Gale Crater, a basin with the area of Lake Ontario. Its goal is to explore the ancient rocks of Mars for organic molecules and the remains of watery environments that could have provided a habitat for life. It will do that by climbing a

mountain that rises from the centre of the crater: Aeolis Mons — informally dubbed Mount Sharp by mission scientists — which contains 5.5 kilometres of layered rocks representing hundreds of millions of years of Martian history. As the 900-kilogram rover climbs those slopes, on a journey that could take a decade or more, it will carry not only the most extensive suite of instruments ever sent to Mars, but also the hopes and dreams of engineers and explorers who see the mission as a prelude to an eventual human presence on the planet.

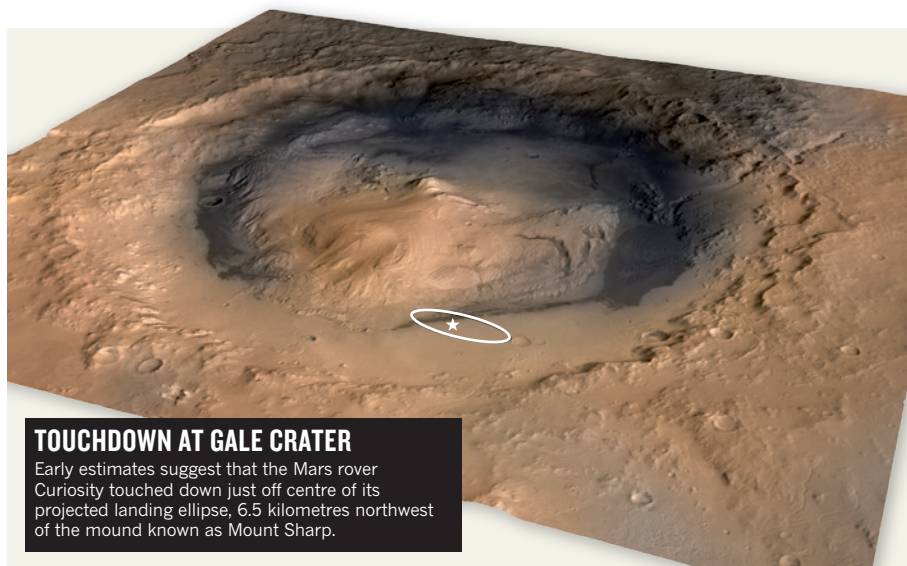
"The wheels of Curiosity have begun to blaze a trail for human footprints on the surface of Mars," said NASA administrator Charles Bolden at an emotional briefing at the Jet Propulsion Laboratory (JPL) in Pasadena, California, shortly after the landing. For now, the 400-strong rover science team must work to ensure that their US\$2.5-billion mission exceeds the achievements of all previous excursions to the red planet.

Complicated technology was needed to drop the rover precisely between Mount Sharp and the walls of Gale Crater. Over 7 minutes, a combination of a heat shield, a parachute, retrorockets and a 'sky crane' decelerated the rover from 5,900 metres per second to less than 1 metre per second and set it down gently on the surface (see *Nature* 488, 16–17; 2012).

"Touchdown confirmed," said Allen Chen, the JPL's operations lead for entry, descent and landing, whose voice remained calm and steady throughout the tense sequence. His words triggered hugs, high-fives and tears of relief. Moments later, the first pictures from the rover's front and rear hazard-avoidance cameras were relayed to Earth by Mars Odyssey, an 11-year-old orbiter that was passing over the crater.

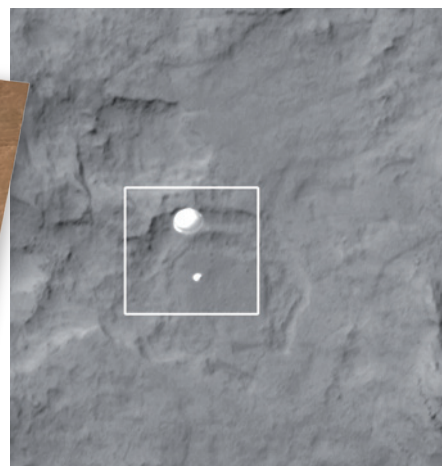
Doug McCuistion, director of the Mars exploration programme at NASA headquarters in Washington DC, doesn't want this new landing capability, developed over the course of a decade, to go to waste. A rover like Curiosity could be built more cheaply and quickly now, he says, because he directed the JPL to "treat this like we're going to build them again and again and again". If the mission were to be repeated, the rover might cost 500 million dollars less. But that is still beyond the budget of NASA's planetary science programme, which has no further Mars landings on the books for ▶

➔ **NATURE.COM**
For more on the Curiosity mission, visit:
nature.com/curiosity



TOUCHDOWN AT GALE CRATER

Early estimates suggest that the Mars rover Curiosity touched down just off centre of its projected landing ellipse, 6.5 kilometres northwest of the mound known as Mount Sharp.



Passing above the rover, NASA's Mars Reconnaissance Orbiter captured a photo of Curiosity parachuting down to the surface.

► now. However, the programme does have an opportunity for a Mars mission in 2018, costing between \$700 million and \$800 million; later this month, NASA science chief John Grunsfeld is scheduled to reveal the results of a study exploring the best use of this window.

While the descent team basked in the post-landing limelight, mission engineers set about testing the car-sized rover. During the vehicle's first hours on Mars, data arrived in a relative trickle. The following day, engineers commanded the rover to deploy a communications antenna that will increase the data rate. The unstowing of the rover's mast, which contains several cameras, is planned for the day after that, and should unleash a cascade of colour photos of the landing site later this week. It will be a further several days before a drive is attempted.

The rover is starting its exploration essentially where mission planners had hoped. The

day after the landing, Mike Malin, president of Malin Space Science Systems in San Diego, California, and principal investigator for a camera on the belly of the rover, unveiled a dramatic video showing the final 150 seconds of the rover's descent. By cross-checking those images with high-resolution photos from the Mars Reconnaissance Orbiter — which caught the spacecraft in the act of descending beneath its parachute — Malin was able to pinpoint the rover on a barren plain just 6.5 kilometres from Mount Sharp (see 'Touchdown at Gale Crater').

Now, project scientist John Grotzinger, a geologist at the California Institute of Technology in Pasadena, needs to decide on the direction of the rover's first foray. Should he turn Curiosity away from Mount Sharp and drive it towards an intriguing fan of material at the crater's rim? The feature, called an alluvial

fan, is thought to have been formed by water that swept sediments over the lip of the basin. An instrument on Odyssey has detected that materials in the fan retain heat longer than the surrounding soil in the cold Martian night. Grotzinger says that this could indicate that the materials in the fan are firmer and more consolidated than the rest. "That implicates water as one way to cement them together," he says.

Alternatively, Curiosity could travel in the opposite direction: towards the base of Mount Sharp, where rock beds contain water-altered clays and sulphates. Whichever direction he chooses, Grotzinger wants to explore the rocks near the landing site to learn how layers of sediment from the base of Mount Sharp interweave with layers from the alluvial fan, and find out which was laid down first. Answers might lie in small craters or other features that cut through the rock layers near the rover, so the mission team will create a route based on those objects. "We're going to try to string together as many pearls as we think we can identify from orbit, and then explore them as we drive along," says Grotzinger.

Curiosity is more powerful than NASA's previous Mars rovers, Spirit and Opportunity, and more efficient at exploring a three-dimensional area. And the strata of the crater and mountain offer an abundant view of the fourth dimension — time. Grotzinger notes that although Spirit and Opportunity have travelled more than 42 kilometres between them since they landed in 2004, they have crossed only tens of metres of strata. At Mount Sharp, Curiosity has 5,500 metres of strata to traverse.

Curiosity itself has time on its side. Its nuclear-powered energy source could sustain it for many years beyond its nominal two-year lifetime. So as tempting as it is to turn Curiosity into a mountain climber immediately, Grotzinger is happy to be patient. "If it takes a year to get there," he says, "that's okay." ■ SEE

EDITORIAL P.129



The Curiosity mission team celebrated after the rover touched down according to plan.

BIOMEDICINE

Stem-cell pioneer banks on future therapies

Japanese researcher plans cache of induced stem cells to supply clinical trials.

BY DAVID CYRANOSKI

Progress toward stem-cell therapies has been frustratingly slow, delayed by research challenges, ethical and legal barriers and corporate jitters. Now, stem-cell pioneer Shinya Yamanaka of Kyoto University in Japan plans to jump-start the field by building up a bank of stem cells for therapeutic use. The bank would store dozens of lines of induced pluripotent stem (iPS) cells, putting Japan in an unfamiliar position: at the forefront of efforts to introduce a pioneering biomedical technology.

A long-held dream of Yamanaka's, the iPS Cell Stock project received a boost last month, when a Japanese health-ministry committee decided to allow the creation of cell lines from the thousands of samples of fetal umbilical-cord blood held around the country. Yamanaka's plan to store the cells for use in medicine is "a bold move", says George Daley, a stem-cell biologist at Harvard Medical School in Boston, Massachusetts. But some researchers question whether iPS cells are ready for the clinic.

Yamanaka was the first researcher to show, in 2006, that mature mouse skin cells could be prodded into reverting to stem cells¹ capable of forming all bodily tissues. The experiment, which he repeated² with human cells in 2007, could bypass ethical issues associated with stem cells derived from embryos, and the cells could be tailor-made to match each patient, thereby avoiding rejection by the immune system.

Japan is pumping tens of millions of dollars every year into eight long-term projects to translate iPS cell therapies to the clinic, including a US\$2.5-million-per-year effort to relieve Parkinson's disease at Kyoto University's Center for iPS Cell Research and Application (CiRA), which Yamanaka directs. That programme is at least three years away from clinical trials. The first human clinical trials using iPS cells, an effort to repair diseased retinas, are planned for next year at the RIKEN Center for Developmental Biology in Kobe.

Those trials will not use cells from Yamanaka's Stock. But if they or any other iPS cell trials

succeed, demand for the cells will explode, creating a supply challenge. Deriving and testing iPS cells tailored to individual patients could take six months for each cell line and cost tens of thousands of dollars.

Yamanaka's plan is to create, by 2020, a standard array of 75 iPS cell lines that are a

leukaemia. Each bank will determine for itself whether further consent is needed.

Yamanaka has already built a cell-processing facility on the second floor of CiRA and is now applying for ethics approval from Kyoto University to create the stock. Takafumi Kimura, a CiRA biologist and head of the project's HLA analysis unit, says that the team hopes to derive the first line, carrying a set of HLA proteins that matches that of 8% of Japan's population, by next March.

Yamanaka's project has an advantage in that genetic diversity in Japan is relatively low; elsewhere, therapeutic banks would have to be larger and costlier. Most iPS banks outside Japan specialize in cells from people with diseases, for use in research rather than treatment. The California Institute for Regenerative Medicine (CIRM) in San Francisco, for example, plans to bank some 3,000 cell lines for distribution to researchers.

Alan Trounson, president of CIRM, says that unresolved research questions about iPS cells make it "premature" to begin therapeutic trials. "We don't have complete pictures of how good they would be," he says, noting that such cells accumulate mutations and other defects as they are produced from differentiated cells. Irving Weissman, a stem-cell biologist at Stanford University in California, warns that iPS cells derived from blood cells have been shown to form tumours³.

Kimura says that the answer is to carefully avoid the white blood cells that cause tumours when deriving the cell lines, and he stresses that all safety concerns will be addressed. "We're building a national resource. It has to be safe and have the confidence of the people."

Daley, who last month toured CiRA's facility, calls it "nothing short of spectacular, pristine, perfect". He agrees that proving the safety of the cells will be tough, but he is enthusiastic about the effort. "It's clear they're readying themselves for a big project," he says. ■

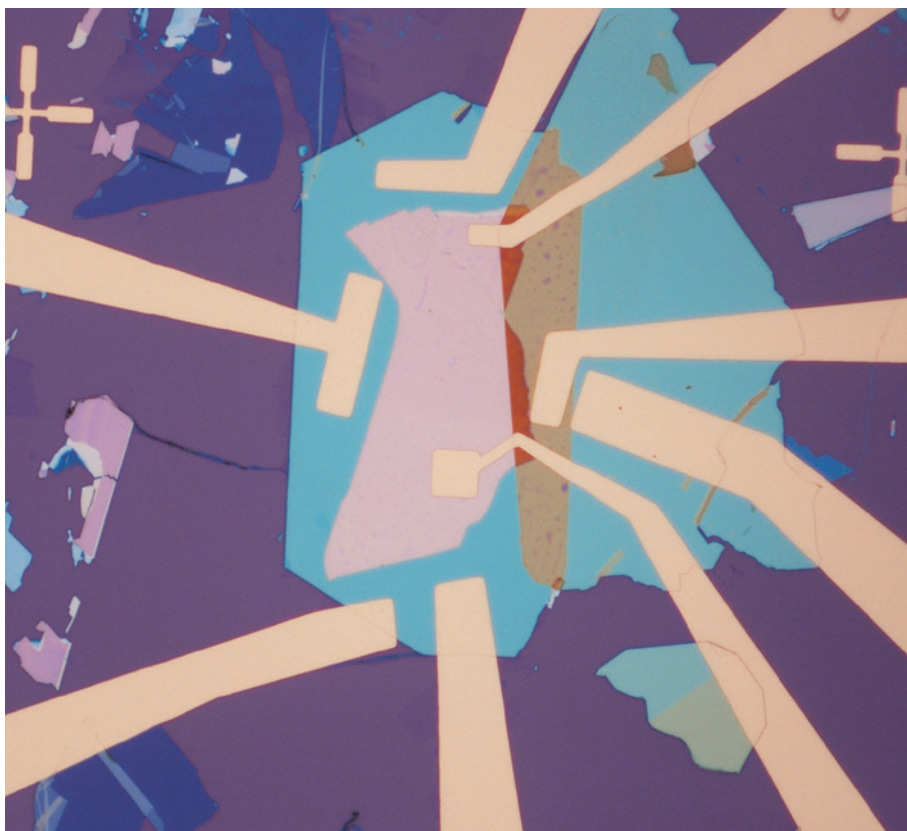


Shinya Yamanaka aims to produce cell lines from fetal blood cells.

good enough match to be tolerated by 80% of the population. To do that, Yamanaka needs to find donors who have two identical copies of each of three key genes that code for immune-related cell-surface proteins called human leukocyte antigens (HLAs). He calculates that he will have to sift through samples from some 64,000 people to find 75 suitable donors.

Using blood from Japan's eight cord-blood banks will make that easier. The banks hold some 29,000 samples, all HLA-characterized, and Yamanaka is negotiating to gain access to those that prove unusable for other medical procedures. One issue remains unresolved: whether the banks need to seek further informed consent from donors, most of whom gave the blood under the understanding that it would be used for treating or studying

1. Takahashi, K. & Yamanaka, S. *Cell* **126**, 663–676 (2006).
2. Takahashi, K. *et al.* *Cell* **131**, 861–872 (2007).
3. Serwold, T. *et al.* *Proc. Natl Acad. Sci. USA* **107**, 8939–8943 (2010).



Layering graphene with other crystalline materials can create electronic devices such as transistors.

TECHNOLOGY

Britain's big bet on graphene

Manchester institute will focus on commercial applications of atom-thick carbon sheets.

BY GEOFF BRUMFIEL IN MANCHESTER

Kostya Novoselov has the tour down pat. After a friendly introduction, visitors are whisked to a clean room so that they can repeat the experiment that helped to win him a share of the Nobel Prize in Physics in 2010. The important bit can be done in seconds: press some sticky tape onto a chunk of graphite, then press it again onto an ultraclean silicon wafer. Peel it off, and some of the silver flakes dotting the wafer's surface are atom-thick sheets of honeycombed carbon known as graphene.

The material has had a meteoric rise since Novoselov, his fellow Nobel laureate Andre Geim and their team at the University of Manchester, UK, reported this deceptively simple way of making graphene¹. Hundreds

of groups around the world have investigated its remarkable range of properties. It is highly conductive, and exhibits a variety of quantum-mechanical behaviours that had previously been seen only in more complex materials. It is thin and flexible, and its electrical and mechanical properties change in response to its surroundings. These characteristics and others suggest various electronics applications, including touch screens, sensors and frequency generators.

Now, the UK government is hoping that Novoselov and Geim can make money from graphene. In February, the UK Engineering and Physical Sciences Research Council (EPSRC) announced £38 million (US\$59 million) in funding for a National Graphene Institute at the university. Scheduled to open in 2015, the centre will be a hub for

translating basic research into industrial applications. Researchers at Manchester will mingle with industrial scientists loaned by domestic and overseas technology firms. Spin-off companies will flourish in off-campus research parks, sparking a technology revolution in a city that was once at the centre of the Industrial Revolution. That's the vision, at least.

The initiative has its share of critics. Some academics in the United Kingdom say that the government wants to turn campuses into money-making enterprises. Elsewhere, it is not difficult to find researchers who say that graphene's commercial value may be overhyped. Many doubt that a single material, however promising, can revive Britain's atrophied electronics industry.

Novoselov acknowledges these arguments, but thinks that the investment will pay off. "These millions will come back quite soon," he says. It will certainly mean an expansion of his own research enterprise, which is currently looking at how to combine graphene with other two-dimensional materials. It will also allow companies large and small to come to the university and conduct research that might further their own industrial ambitions.

The University of Manchester is a microcosm of a global boom in graphene research (see 'Graphene goes global'). More than 20 academics from its chemistry, biology, materials science and engineering departments participate in weekly meetings about the material. The discussions are not only about electronics: some colleagues are studying graphene for use in biosensors, and others want to incorporate it into advanced materials.

Demand for graphene from other departments has been so high that the group cannot keep up, so Branson Belle, one of Novoselov's postdocs, launched a start-up firm in May to supply the rest of the university. Novoselov and Geim's core group is growing too, and lab space is getting tight.

Novoselov and Geim were already asking the university for more room last year, and administrators were thinking about how to capitalize on the graphene explosion, says Nancy Rothwell, president and vice-chancellor of the University of Manchester. The pair, together with the university and the city council, submitted a proposal to the EPSRC, which awarded the university £26 million to build the graphene institute, along with £12 million for equipment.

The funding comes at a time of austerity for UK scientists, and it has raised eyebrows in other parts of the physical-sciences community. "The institute looks great and will certainly produce outstanding science," says

Pascal André, who works on nanomaterials at the University of St Andrews, UK. But he questions the wisdom of plunging £38 million

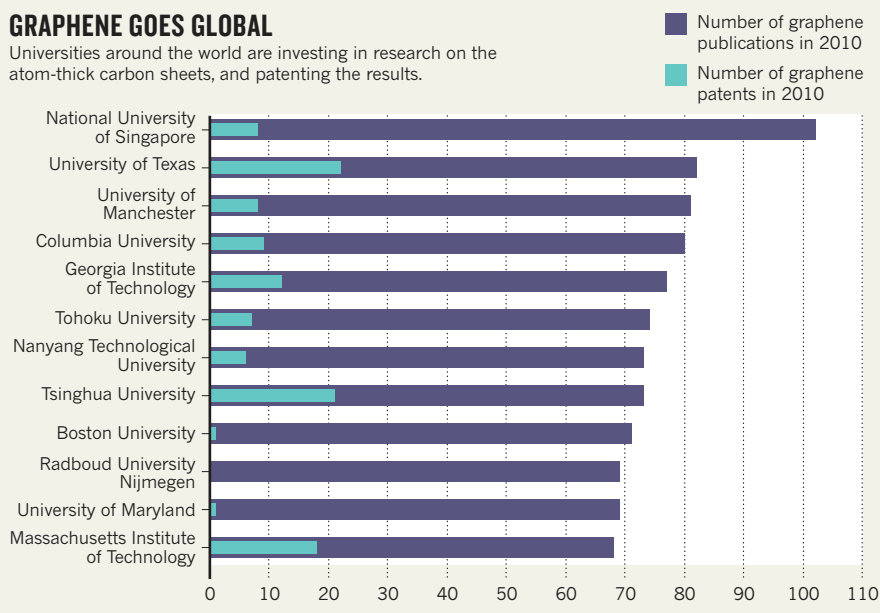
➔ **NATURE.COM**

For more, see the *Nature Outlook* on graphene:

go.nature.com/otkrbs

GRAPHENE GOES GLOBAL

Universities around the world are investing in research on the atom-thick carbon sheets, and patenting the results.



into a single subject and location when the EPSRC is slashing research budgets elsewhere. “Will the whole UK innovation of the next 5–10 years be solely based around graphene?” André asks.

Novoselov says that he agreed to the institute only after he was assured that the money would not come at the expense of research grants.

Whether the government’s gambit will help the United Kingdom to compete in the global electronics market is uncertain. Britain lags behind the rest of the world in graphene patents, according to Quentin Tannock, the chairman of Cambridge IP, an independent patent consulting firm in Cambridge, UK. The country is competing with others such as South Korea, a major producer of consumer electronics, which already has a US\$20-million, five-year programme to develop graphene-based display panels and other devices, according to Byung Hee Hong, a graphene researcher at Seoul National University. Hong’s group has been perfecting methods for producing sheets of graphene on an industrial scale. “I’m now working with seven different companies,” he says. But Hong adds that the strong commercial interest in Seoul may actually bode well for the new Manchester centre. “Korean companies are not working only in Korea,” he says.

Even if companies are curious, graphene may still flounder as a commercial product, cautions Phaeton Avouris, a materials scientist at the IBM T. J. Watson Research Center in Yorktown Heights, New York. “There has been this circulating myth that

graphene will replace silicon,” Avouris says. In fact, the material is not a semiconductor and lacks the necessary bandgap that would allow it to serve as a transistor — the basic element of all electronics — on its own. Avouris thinks that the material could find a use in niche markets such as high-frequency electronic devices, but he questions whether it will ever hit the big time.

Novoselov does not disagree. But, he adds, there have been enough suggested uses for graphene in different areas to make him want to look at other applications. “What I know for sure is that if we are not going to work on this, it will definitely not happen,” he says.

Novoselov has thrown himself into the project to build the graphene institute, meeting with architects and possible industrial partners on a weekly basis. Geim, who declined to be interviewed, is heavily involved too. Their group, however, is still at the cutting edge of graphene research, and is investigating the properties of layered graphene as well as other two-dimensional crystals such as boron nitride, molybdenum disulphide and niobium diselenide. Like graphene alone, the layered systems display exotic quantum behaviours and could have various applications. For example, in February, the group reported building a transistor by sandwiching boron nitride between two graphene sheets. When a voltage was applied, electrons tunneled from one graphene sheet to the other, through the boron nitride barrier².

“We’ve been doing good science for some time and we will do it for the years to come,” says Novoselov. But he is quick to add that now is the right time to make that science pay. “Our government is right: if we have this chance, we should take it.” ■

1. Novoselov, K. S. *et al. Science* **306**, 666–669 (2004).
2. Britnell, L. *et al. Science* **335**, 947–950 (2012).

“Will the whole UK innovation of the next 5–10 years be solely based around graphene?”



Droughts are becoming more frequent in parts of the United States.

CLIMATE CHANGE

Heatwaves blamed on global warming

Unusually high frequency points to human influence.

BY JEFF TOLLEFSON

NASA climatologist James Hansen made headlines during the US heatwave of 1988, declaring in testimony to Congress and during interviews on prime-time television that a build-up of greenhouse gases was increasing the probability of weather extremes. Now, as much of the United States sizzles through another torrid summer and the Midwest endures a historic drought, Hansen, director of NASA's Goddard Institute for Space Studies in New York, claims that the future he predicted has arrived.

"The climate dice are now loaded to a degree that a perceptive person old enough to

remember the climate of 1951–1980 should recognize the existence of climate change, especially in summer," he and his colleagues write in a paper entitled 'Perceptions of Climate Change' published on 6 August. Just days earlier, on 1 August, Republican senators had challenged mainstream climate scientists over the existence of anthropogenic global warming at a hearing in Washington DC, underscoring the stubborn political divide over climate policy. Just as he did 24 years ago, Hansen has plunged into the debate, pre-empting the publication of his study with an opinion article in *The Washington Post*².

Hansen's team used seasonal temperature records for 1951–80, a period of relatively

stable climate, as a baseline, then analysed the frequency and scale of subsequent temperature anomalies. On average, the team concludes, the globe has warmed by only about 0.5–0.6°C since that time, but the shift has had a significant impact on many parts of the world (see 'What a scorcher').

Extremely hot summers — classified as about 3.5°C warmer than average — have affected about 10% of the world's land since 2006, an order of magnitude higher than during the period from 1951 to 1980.

The study is not the first to show a link between global warming and extreme weather³, but it goes well beyond its predecessors, concluding that greenhouse gases alone are responsible for the hot summers and heatwaves. "The likelihood that these events would have occurred without global warming is minuscule," Hansen says.

A poll by researchers at Yale University in New Haven, Connecticut, and George Mason University in Fairfax, Virginia, suggests that most people in the United States accept the link between hot weather and global warming⁴. But Hansen's assertion is running into some heavy weather among scientists.

Martin Hoerling, a meteorologist at the National Oceanic and Atmospheric Administration in Boulder, Colorado, calls Hansen's paper an "extended Op-Ed piece", arguing that the broader climate record does not support the link to individual heatwaves. Last year, Hoerling co-authored a paper⁵ suggesting that the 2010 drought in Russia was so far outside the realm of normal weather that the small rise in global temperatures could not account for it. He says that natural variability can explain most extremes, and that global warming merely enhances them.

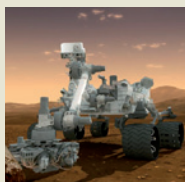
Hansen notes that his study is purely statistical and does not try to explain how climate change could cause extremely hot summers. Kevin Trenberth, a climatologist at the US National Center for Atmospheric Research (NCAR) in Boulder, says that Hansen's statistics are illustrative of a trend that should help people to understand global warming and the profound effect humans have had on the climate system. "It is never due to humans alone, nor is it ever, these days, just natural variability."

In a paper to be published in the *Journal of Geophysical Research*⁶, Trenberth and a ▶



**MORE
ONLINE**

LIVE BLOG

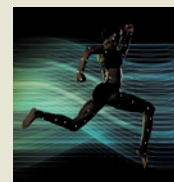


Relive the moment that NASA's Curiosity rover landed on Mars
go.nature.com/onaw4q

NEWS

- Guinea pig hearts beat with human cells go.nature.com/8li8uu
- Scientists record signal as distant black hole consumes star go.nature.com/hwk2wh
- Pregnancy alters resident gut microbes go.nature.com/4v7cht

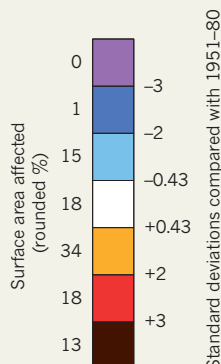
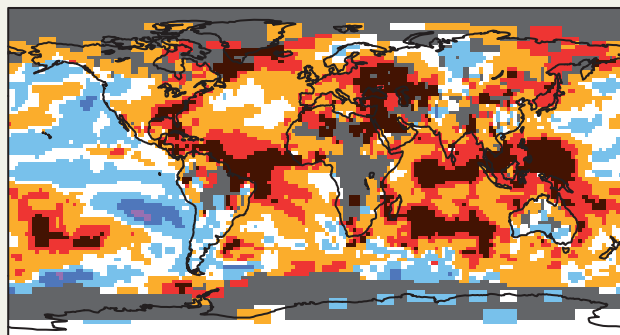
OLYMPICS



Nature takes a look at how research has shaped the events at London 2012
go.nature.com/thgovl

WHAT A SCORCHER

Between June and August 2010, extremely high temperatures (brown) hit about 13% of Earth's surface, an area roughly ten times greater than in 1951–80.



► team of researchers investigate the physical mechanisms that drove some extreme weather events in 2010. Using a climate model developed at the NCAR, the team investigated

links between a pair of El Niño and La Niña events (in which warm or cold surface waters, respectively, built up in the eastern Pacific Ocean) and weather events such as stronger

monsoons in Asia and droughts in Russia and the Amazon. Although he thinks that global warming could have a role in such extreme events, Trenberth says that climate models have not yet been able to tease out the details.

"Models have a hard time doing extremes well," Trenberth says. But because of limited data sets for extreme weather and inadequate climate models, he worries that some people could draw the wrong conclusion: "that there is no human influence". ■

1. Hansen, J., Sato, M. & Ruedy, R. *Proc. Natl. Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1205276109> (2012).
2. Hansen, J. E. *The Washington Post* (5 August 2012).
3. Schiermeier, Q. *Nature* **477**, 148–149 (2011).
4. Leiserowitz, A., Maibach, E., Roser-Renouf, C. & Hmielowski, J. *Global Warming's Six Americas*, March 2012 & Nov. 2011 (Yale Univ. & George Mason Univ, 2012).
5. Dole, R. *et al. Geophys. Res. Lett.* **38**, L06702 (2011).
6. Trenberth, K. E. & Fasullo, J. J. *Geophys. Res.* <http://dx.doi.org/10.1029/2012JD018020> (in the press).

AWARDS

Physics prize dwarfs all others

Theorists left reeling after billionaire reveals massive prize.

BY GEOFF BRUMFIEL

When Alan Guth received an e-mail from a colleague asking if he could discuss a new annual prize in physics, he recalls, "I thought I was being asked to be on an organizing committee".

Instead, the other physicist, Nima Arkani-Hamed at the Institute for Advanced Study in Princeton, New Jersey, told Guth that they were among the winners of a US\$3-million award. Guth, a theorist at the Massachusetts Institute of Technology in Cambridge who introduced the idea of cosmic 'inflation', assumed that the prize would be split. But Arkani-Hamed surprised him again: they had won \$3 million each. "Do you mean that I just won \$3 million and that you just won \$3 million?" Guth asked, incredulously. "He said 'yes'."

"At that point I just kind of fell off my seat," Guth says.

Last week's public announcement of the prize, which was awarded to nine physicists, was just as sudden (see go.nature.com/mwaays). On 31 July, two weeks after Guth's unexpected conversation, stories appeared in *The New York Times* and *The Guardian* unveiling the \$27-million Fundamental Physics

Prize, which dwarfs all other prizes in science. Shortly afterwards, a bare-bones website appeared. Under the heading 'board', it listed just two people: Steven Weinberg, a Nobel-prizewinning theorist at the University of Texas at Austin, and a man named Yuri Milner.

It was Milner, a 50-year-old Russian Internet entrepreneur, who founded the prize and chose the first winners (Weinberg was appointed to the board only after Guth and the others were notified). In his youth, Milner spent a decade studying theoretical physics in the Soviet Union before abandoning his PhD to move into the private sector. His early investments included a macaroni factory and a Russian e-mail portal, but Milner's fortune was sealed in 2009 when his firm bought a \$200-million stake in the social-media site Facebook. In May, Facebook began trading publicly, and Milner's company made an estimated \$1.7 billion. Today, his investment funds are valued at \$12 billion, and his personal wealth at around \$1 billion.

Milner never forgot his early years working on quantum chromodynamics, a theoretical framework that describes the interactions of quarks

and gluons. For years he had toyed with the idea of creating a prize and, in recent months, he decided to act on it, choosing nine theorists who have never won Nobel prizes, but whose work he considers to be groundbreaking. "It was clear that he'd done a lot of homework," Arkani-Hamed says. "He knew an impressive amount about what was going on."

"The intention was to say that science is as important as shares trading on Wall Street," Milner told *Nature*. The prize money comes with no strings attached, although Milner hopes that the theorists will contribute to a new lecture series for the public. Milner also plans to create an annual \$100,000 New Horizons prize for young researchers and an ad hoc version of the Fundamental Physics Prize that can be won at any time "in exceptional cases", the website says. According to the rules of the prizes, anyone can be nominated, and future prizewinners will be selected by a committee of all previous ones.

"He's recognizing that there are some very smart people who have done some very clever things," says George Smoot, a physicist at the University of Paris Diderot who shared the 2006 Nobel Prize in Physics. But, he adds, giving such a large prize to top theorists, many of whom are late in their career, will not revolutionize theoretical physics. Smoot, who donated his portion of the Nobel prize to charity, says that he would have liked to have seen Milner put half of the \$27 million into fellowships for young researchers. Arkani-Hamed says that Milner is aware of the criticisms but felt that other philanthropists have created fellowships and institutes, and he wanted the new prize to stand apart.

For his part, Guth says that he hasn't yet decided what to do with the money, which was deposited directly into his bank account while he and his wife were away at a local Shakespeare festival. "We're still kind of in shock," he says. ■

► **NATURE.COM**
To read about
another major prize,
see:
go.nature.com/sjsxr6



BUCKLING DOWN

Mechanical instability is usually a problem that engineers try to avoid. But now some are using it to fold, stretch and crumple materials in remarkable ways.

Katia Bertoldi is talking fast. She has only 12 minutes to present her work in the burgeoning field of 'extreme mechanics'. But first, the Harvard University engineer smiles at the physicists gathered in Boston at the March 2012 meeting of the American Physical Society. She has to show them what she found in a toy shop.

Projected onto the screen, the Hoberman Twist-O looks like a hollow football made of garishly coloured plastic links. Twist it just so, however, and hinges between the links allow it to collapse into a ball a fraction of its original size. Twist it the other way, and it springs back open. Bertoldi explains that the Twist-O inspired her group to create a spherical device that collapses and re-expands, not with hinges but through mechanical instabilities: carefully designed weak spots that behave in a predictable way. Applications might include lightweight, self-assembling portable shelters or nanometre-scale drug-delivery capsules that would expand and release their cargo only after they had passed through the bloodstream and reached their target.

The challenge, Bertoldi says, is to figure out the exact instabilities a

BY KIM KRIEGER

structure needs to achieve its desired behaviour. She quickly describes the necessary geometry and runs down a list of constraints. There are

just 25 shapes that satisfy all the requirements, she explains, glossing over the months of computation it took to solve the problem. Then she starts a video to show the assembled through the design that her team has come up with.

An image of a rubbery chartreuse ball with 24 carefully spaced round dimples (**pictured**) materializes on the screen. The test begins and the ball slowly collapses, each dimple squeezing shut as the structure twists into a smaller version of itself. There is a moment of silence, then everyone in the room begins to clap.

Student engineers have always been taught that mechanical instabilities are a problem to avoid. Such instabilities can quickly lead to structural failures — the collapse of a weight-bearing pillar, the crumpling of a flat steel plate or the buckling of a metal shell. From failures come disasters, such as the Second World War Liberty Ships that broke up while at sea. And the devilishly complex mathematical analysis of

JONGMIN SHIM, KATIA BERTOLDI AND PEDRO REIS

'Buckliballs' collapse and re-expand owing to careful placement of mechanical instabilities.

buckling structures ground to a halt in the late nineteenth century, because it was unworkable with the methods then available. During the past half-decade or so, however, a new generation of physicists and engineers has begun to embrace instability. These researchers have been inspired, in part, by advances in geometry and nonlinear mathematics that have allowed them to progress where their forebears could not. They have already, for example, devised a theory for why cabbage leaves and torn plastic rubbish bags ripple¹; calculated the patterns of wrinkles in fabric and crumples in paper²; and accounted for the way coils and loops develop in the guts of vertebrate embryos³.

On the practical side, one source of inspiration has been the widespread availability of flexible polymers and silicone materials, as epitomized by the vast selection of soft yet tough covers for smart phones. Such materials make it possible to imagine electronics, robots, tools and vehicles whose structures can radically deform yet still recover their original shapes.

The resulting extreme-mechanics movement has grown rapidly. The first three conference sessions to bear the name, totalling fewer than 40 presentations, were held during the March 2010 meeting of the American Physical Society (APS). Just two years later, Bertoldi's talk on collapsible spheres was one of 111 presentations on extreme mechanics spread across 8 sessions. Hundreds of researchers are now active in the field worldwide. In spring 2011, the US National Science Foundation announced an opportunity for substantial funding in the field: it would allot up to US\$2 million over four years to projects in Origami Design for Integration of Self-assembling Systems for Engineering Innovation (ODISSEI). The foundation expects to announce the awards this month.

FOLD EVERYTHING

"It was as if they wrote the solicitation just for us," says Christian Santangelo, a physicist at the University of Massachusetts Amherst and a co-principal investigator on an ODISSEI proposal. Along with two origami artists and an expert in origami mathematics, Santangelo and his colleague Ryan Hayward, a chemical engineer at Amherst who specializes in polymers, are proposing a new kind of three-dimensional (3D) printer. Instead of slowly building an object with layers and layers of polymer, as current 3D printers do, they would print a flat polymer sheet with a two-dimensional (2D) origami-like pattern, then force it to fold into a close approximation of the desired 3D object.

One part of the project will involve refining a computer program developed by one of the origami artists, physicist Robert Lang of Alamo, California. Given a desired shape, the program will generate a diagram of the required fold pattern. At present, the program states only whether a particular fold on a sheet should be convex (in origami terms, a mountain fold) or concave (a valley fold). The user still has to devise a sequence of manipulations that can achieve those folds and create the figure. But the kinds of folding required by the project could quickly reach such levels of complexity that a human solution would be impossible. What the researchers foresee instead is a completely automated process in which a 2D sheet is inscribed with a computer-generated pattern of instabilities, and then folds correctly in one smooth, coordinated motion.

Unfortunately, in Hayward's experiments so far, the folds just buckle into mountains or valleys at random. A potential solution may lie in the wavy-leaf-edge phenomenon, says Santangelo. If cells along the edge of a growing plant leaf multiply faster than those in the interior, he explains, they run out of room to lie smoothly in a plane, and the leaf edge is forced to ripple to accommodate them. If he and Hayward can

work out how to make the polymer sheet swell in a way that varies from point to point, they could produce a complex pattern of rippling and curling that would help to control the sheet's folding.

Just as natural phenomena can inform extreme mechanics, the discipline's researchers can also use their knowledge to explain peculiarities in natural structures. "This field is filled with small secrets," says Pedro Reis, an engineer at the Massachusetts Institute of Technology in Cambridge and one of the leaders of the movement.

Take the mechanisms at work in a grain of pollen, for example. Reis pulls out a small torpedo-like shape made of a light green, rubbery material. He stretches it, then crushes it in his fist. A pollen grain undergoes torture, he says: it gets wet, swells, dries out and is crushed, so plants have evolved strategies such as built-in soft spots to help their pollen to avoid damage. Reis pokes a dimple in the torpedo. Scientists can learn from this that a shell with a soft spot is more resistant to failure than one that is completely rigid, he says. "We are trying to learn from nature. How it evolved to deal with these problems for which we have no intuition is very inspiring."

Reis sets the torpedo aside and, twisting an imaginary string between his fingers, moves on to the subject of the sea-floor cables that carry Internet traffic around the globe. Mechanical instabilities can cause complex behaviour in these structures, too, he says. Lay too much cable, and it will curl and kink, resulting in poor signal quality. Lay too little, and the line will be tense and vulnerable to snapping. When a ship dragging its anchor sliced through one such cable in February, six countries in East Africa lost Internet connectivity. Although a better understanding of long, thin objects' instabilities wouldn't have prevented the accident, Reis says, it might have made for a cable more resilient to breakage. An improved theory of cables could also be of great use to fields ranging from the oil industry to the mechanics of DNA.

In an effort to turn such insights into engineered structures, Reis's lab is competing for one of the ODISSEI grants in partnership with Bertoldi and several others. One of their goals is to compile a lexicon of shapes: polyhedra that will buckle, bend, stretch, collapse and expand in predictable ways in response to specific stimuli. The shapes are based on spherical shells with holes cut in them. But surrounding the holes are ligaments designed to contain weak spots that buckle when stressed. In principle, these buckling polyhedra could be made in a range of sizes, whether it's a nanometre-scale sphere designed for drug delivery or a retractable roof for an athletics stadium. The team calls the resulting structures 'buckliballs' — the subject of Bertoldi's presentation at this year's APS meeting.

Buckliballs, with their geometric design and near-magical behaviour, encapsulate the cleverness and beauty for which extreme-mechanics researchers yearn. Looking ahead, the more theoretically minded foresee a new set of general rules that could describe the behaviour of any flexible solid as it crumples. Meanwhile, those with an engineering bent imagine robots with appendages that can transform into tools or squeeze, octopus-like, through tiny spaces; backpacks that expand into tents; and mobile phones that users can roll up and stick behind their ears like a pencil. They see a whole realm of devices that transmute failure into function. That could all be years in the making — but Bertoldi and her rapt audience already see far more in this field than engineers' toys. ■

Kim Krieger is a freelancer writer in Norwalk, Connecticut.

1. Sharon, E., Marder, M. & Swinney, H. *Am. Sci.* **92**, 254–261 (2004).
2. Vandeparre, H. *et al. Phys. Rev. Lett.* **106**, 224301 (2011).
3. Savin, T. *et al. Nature* **476**, 57–62 (2011).



OPEN AMBITION

Jay Bradner believes that cancer can be defeated through control of epigenetics — and he is not shy about spreading the word.

BY AMY MAXMEN

Jay Bradner has a knack for getting the word out online. You can follow him on Twitter; you can become one of more than 400,000 online viewers of the TEDx talk he gave in Boston, Massachusetts, last year; you can see the three-dimensional structure of a cancer-drug prototype created in his laboratory and you can e-mail him to request a sample of the compound.

Bradner, a physician and chemical biologist at the Dana-Farber Cancer Institute in Boston, makes defeating cancer sound easy — one just has to play tricks on its memory. “With all the things cancer is trying to do to kill our patient, how does it remember it is cancer?” he asked his rapt TEDx audience. Bradner says that the answer lies in epigenetics, the programmes that manage the genome.

DNA serves as the basic blueprint for all cellular activity, and DNA mutations have long been known to have a role in cancer. But much of a cell's identity is determined by modifications to chromatin, which comprises DNA and the proteins that bind and package it. Epigenetic instructions, in the form of chemical marks that cling to chromatin, tell cells how to interpret the underlying genetic sequence, defining a cell's identity as, say, blood or muscle.

Findings over the past ten years have strongly implicated dysregulation of epigenetic instructions in cancer, where growth-driving genes express like crazy and genes that keep cell division in check are silenced. Bradner's aim is to create a drug that can rewrite those instructions so that cancer cells forget what they are and cease their deadly proliferation.

Bradner thinks that this epigenetic approach could strike down one of cancer's most treacherous drivers, the DNA-binding protein Myc. Myc is involved in up to 70% of cancers but is generally considered ‘undruggable’, because the active parts of its structure are not accessible to the kinds of small-molecule drugs that chemists generally create. “Myc is one of those things that people dream of targeting,” says Dash Dhanak, head of cancer epigenetics at GlaxoSmithKline (GSK) in Collegeville, Pennsylvania.

Just as audacious is Bradner's commitment to making his reagents available and his ideas accessible to scientists and laypeople alike, a rare attitude in the highly competitive world of drug discovery. Researchers in Bradner's lab have developed a compound that interferes with Myc by manipulating epigenetic instructions, and he has sent it out to hundreds of collaborators worldwide. “That's not common in practice,” says Bradner, “but from first principles, it's the right thing to do.”

Detractors may scoff at Bradner's flashy approach, but those who have followed his career say that there is substance to go with the style. “Jay has figured out translational science,” says Stuart Schreiber, director of chemical biology at the Broad Institute in Cambridge, Massachusetts, and Bradner's

former postdoctoral adviser. “He's really just good at making important discoveries while staying connected to their clinical potential.”

In 1992, while Bradner was an undergraduate at Harvard University, also in Cambridge, he took a chemistry class taught by Schreiber on small-molecule discovery. By the time he graduated, Bradner knew that he wanted to apply the methods he had learned to cancer-drug development. He headed west to Illinois, to study the disease at the Pritzker School of Medicine at the University of Chicago.

MARKED FOR DEATH

In 1999, Bradner returned to Massachusetts for a clinical residency at Brigham and Women's Hospital in Boston, and in 2004 he joined Schreiber's lab as a postdoctoral researcher. Schreiber's team was researching chemical compounds that override normal epigenetic control of gene expression by modulating chromatin. Such control systems generally involve three types of protein: ‘writers’, ‘readers’ and ‘erasers’ (see ‘Rewriting memory’). Writers attach chemical marks, such as methyl groups (to DNA) or acetyl groups (to the histone proteins that DNA wraps around); readers bind to these marks and influence gene expression; erasers remove the marks. The marks serve as instructions that are passed down as cells divide, providing a sort of cellular memory to ensure that skin cells, for example, beget other skin cells. Epigenetics has become one of the hottest areas of biological research.

Schreiber and his group had long been looking at histone deacetylases (HDACs), eraser proteins that remove acetyl groups from histones. Some chromatin regions in cancer cells contain fewer acetyl groups than those in normal cells, and drugs called HDAC inhibitors increase acetylation. Since 2006, two such drugs

called NUT midline carcinoma (NMC), which typically kills patients within a year of diagnosis. French, an expert on NMC, diagnosed the disease after cardiac surgeons had opened the boy's chest and found a tumour the size of a baseball in his heart. Chemotherapy was not an option, as it would have hampered his recovery from surgery. Bradner and French discussed other treatments.

French had shown in 2003 that NMC is caused by a fusion of two genes¹: *BRD4*, which encodes a reader protein, and a previously unknown gene called *NUT*. This fusion encodes a mutant protein, NUT-BRD4, which seems to act as a reader, spurring errant gene expression and forcing cells to lose their identity and become cancerous. No inhibitors of reader proteins were available then, but French and Bradner knew that vorinostat increased histone acetylation. Perhaps, they thought, if acetylation were increased, NUT-BRD4 would be so busy ‘reading’ elsewhere that it would overlook the region that was causing cells to become cancerous. Bradner calls the concept “the chemical equivalent of a smokescreen”.

It was thin reasoning, but Bradner and French agreed that the emergency at hand warranted an experiment. They tried the drug on cancer cells extracted from the boy, and the cells seemed to forget their cancerous marching orders, reverting from round, proliferating cells into flat, skin-like cells. This gave them the confidence to try treating the boy. The tumour seemed to respond after five weeks, but the toxicity proved too high. “It was taking him four hours to swallow three pills because he kept vomiting them back up,” French says. The boy stopped the treatment and died not long afterwards.

Bradner dwelled on the case, especially on

“PEOPLE LIKE JAY HAVE BEEN TREMENDOUS IN PUSHING FORWARD THE FRONTIERS.”

— vorinostat and romidepsin — have been approved by the US Food and Drug Administration to treat cutaneous T-cell lymphoma, a rare immune-cell cancer that affects the skin. The drugs generated excitement among cancer researchers, but because they block many types of HDAC — in both healthy and cancerous cells — they can be toxic. Several trials for other cancers turned up disappointing results.

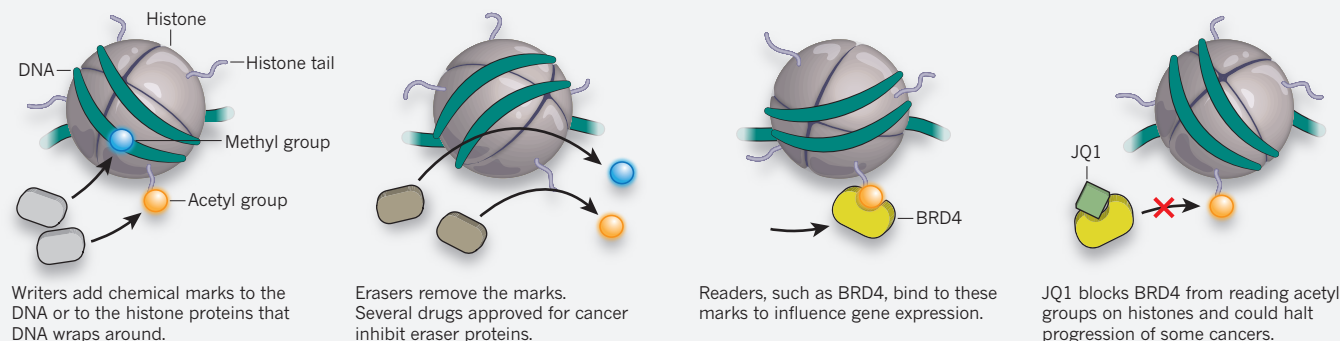
Bradner experienced the let-down of the HDAC inhibitors first-hand in 2008, soon after starting his own lab. Chris French, a pathologist at Brigham and Women's Hospital, consulted with Bradner about a ten-year-old boy he was treating for a rare and aggressive cancer

BRD4, a little-studied protein that now seemed to be capable of causing cancer. Disparate lines of study suggested that BRD4 might be linked to the expression of Myc, a target most drug developers had abandoned. If Bradner could defuse BRD4, perhaps he could bring down one of cancer's most notorious dragons.

In 2009, Bradner happened upon a patent from Mitsubishi Tanabe Pharma in Osaka, Japan, for a diazepam-based compound that inhibited BRD4 by blocking its bromodomain, the region that recognizes acetyl groups on histones. Diazepines on the market, such as the anxiety medication alprazolam (Xanax), have only weak interactions with the bromodomain.

REWRITING MEMORY

Three types of protein — writers, erasers and readers — orchestrate epigenetic instructions, the chemical modifications that regulate gene expression. Researchers think that compounds that manipulate these proteins could stop cancer cells from proliferating.



“You’d be in a coma by the time you inhibited BRD4,” Bradner says. So he began searching for molecules that were similar in structure but more potent. Jun Qi, a researcher in Bradner’s lab, synthesized more than 400 diazepines in search of a candidate.

By the end of 2009, they had one. Named JQ1 after Qi, the compound slips into a groove of BRD4, preventing it from binding to acetylated histones and activating genes. With JQ1, the researchers hoped to tease apart which genes BRD4 switches on — and whether any of them cause cancer.

RESEARCH PROMISE

In March 2010, French introduced Bradner to another patient with NMC, a 29-year-old firefighter from Connecticut. Chemotherapy and romidepsin treatment had failed. Bradner asked the man if he could use his cells for research, pledging that the work would help to find cures. The man agreed. He died that July, but his cells made it possible to test JQ1. The compound stopped the cancer cells from dividing and transformed them into non-cancerous cells, both in culture and in mice².

It takes years for promising molecules to lead to clinically useful drugs. Had drug-company researchers discovered JQ1, a single team would have forged ahead, probably in secret, to develop the compound. But Bradner decided that the quickest path to the clinic was to do the research openly, with as many collaborators as possible. Since January 2011, Bradner’s team has shipped JQ1 to more than 250 labs worldwide. Work on the molecule has produced at least ten publications in top journals.

One of these publications³, a collaboration between Bradner and Constantine Mitsiades, a cancer biologist at Dana-Farber, bolstered BRD4’s connection to Myc. Mitsiades had found that multiple-myeloma cells express high levels of Myc and BRD4, but without a compound to target either of them, it was difficult to learn much more. With Bradner, he found that JQ1 reduced expression of Myc and its target genes and stopped myeloma cells

from dividing in mice³.

Chris Vakoc, a cancer biologist at Cold Spring Harbor Laboratory in New York, was studying the role of BRD4 in leukaemia when he first learned of JQ1. He phoned Bradner immediately. “The next day he sent us a huge amount of the compound,” Vakoc says. In Vakoc’s hands, JQ1 stopped cancer-cell proliferation in mice with leukaemia and significantly extended their lifespans⁴. JQ1 has also been used to study infectious diseases such as those caused by Epstein-Barr virus⁵ and HIV⁶.

“In 20 years, my lab could not accomplish all of the research that has unfolded on JQ1 in one year through this open approach,” says Bradner. With US\$15 million in venture capital from HealthCare Ventures of Cambridge, Massachusetts, Bradner has launched Tensha Therapeutics, a biotechnology company focused on bromodomain inhibition. The Cambridge-based biotech is now screening JQ1 derivatives to learn which are likely to have the fewest side effects — an essential early step in drug development.

Bradner is impatient, however, and is always on the lookout for drugs already in development that would be safe to use in NMC. In 2010, scientists at GSK published a study on an inhibitor of BRD4 and related molecules for treating sepsis⁷. Bradner asked the GSK researchers to try one of their compounds in patients with NMC at Dana-Farber. (Dhanak says that the company had been quietly working on BRD4 inhibitors since French’s *NUT-BRD4* paper in 2003.) GSK agreed.

The first attempt was in March this year, when oncologists at Dana-Farber used the company’s BRD4 inhibitor GSK525762 to treat a 23-year-old engineering graduate student with NMC. The patient died about three weeks into the treatment. French says he suspects that the dose was too low and that combining the drug with HDAC inhibitors or other epigenetic drugs could yield better results. Dhanak says that GSK might test combinations in the near future.

Any clinical trial for NMC is bound to move slowly — the disease is aggressive, and only 90

cases have been diagnosed in the past decade. To raise awareness of NMC and make it easier for patients to enrol in trials, in 2011 Bradner and his colleagues created an international NMC registry (<http://www.nmcregistry.org>). It seems to be working. “We were concerned we’d have a problem finding patients, and now, through social media, the patients are finding us,” Bradner says.

MAKING GOOD

Not all scientists share Bradner’s optimism. “Epigenetics is the new horizon, but when you get down to it, the fact is that people are just mucking around with it and finding interesting effects,” says Gerard Evans, a cancer biologist who studies epigenetic signalling at the University of Cambridge, UK. Epigenetics affects many cellular functions, and researchers are only beginning to learn how it influences cell memory.

Whatever the outcome, some see hopeful signs in Bradner’s open approach. Dhanak says that he welcomes it. “People like Jay have been tremendous in pushing forward the frontiers,” he says.

What Bradner wants most is to fulfil his pledge to the firefighter who died from NMC. “His gift of that rare tumour, given at a time when he was beyond all conceivable treatment, was a powerful experience,” Bradner says. If bromodomain inhibition fails, Bradner will apply the same open-source strategy to another target. “More and more, I feel it is so important to impact patients’ suffering from cancer, and it doesn’t matter whether that’s with our molecules or someone else’s,” he says. ■

Amy Maxmen writes for *Nature* from New York City.

1. French, C. A. *et al. Cancer Res.* **63**, 304–307 (2003).
2. Filippakopoulos, P. *et al. Nature* **468**, 1067–1073 (2010).
3. Delmore, J. E. *et al. Cell* **146**, 904–917 (2011).
4. Zuber, J. *et al. Nature* **478**, 524–528 (2011).
5. Palermo, R. D., Webb, H. M. & West, M. J. *PLoS Pathog.* **7**, e1002334 (2011).
6. Banerjee, C. *et al. J. Leukoc. Biol.* <http://dx.doi.org/10.1189/jlb.0312165> (2012).
7. Nicodeme, E. *et al. Nature* **468**, 1119–1123 (2010).



COMMENT

EPIDEMIOLOGY How disease has clung to the coat-tails of trade down the ages **p.153**

EXHIBITION A variety of human enhancements, from a silver nose to Viagra **p.154**

SCI-FI Sampler of cyberpunk treats features an essay on sitting **p.155**

PHILOSOPHY Science is undergoing a stealth paradigm shift **p.157**

L. YOUNG/PACIFIC RIM CONSERVATION



Two female Laysan albatrosses preen each other at Hawaii's Kaena Point Natural Area Reserve.

Let's talk about sex

The media loves to sensationalize research on animal sexual behaviour — so be careful what you say, warn
Andrew B. Barron and Mark J. F. Brown.

Evolutionary biologists and neuroscientists are often interested in variations in animal sexual behaviour — and particularly relationships between animals of the same sex. How did such traits evolve, and what are their functions and biological bases? Although worthwhile, such research can fuel some of the most licentious scientific reporting in both the mainstream media and specialized publications — titillating prose that wildly misinterprets the research and its implications for human behaviour.

Of course, journalists should be allowed to use crafty wordsmanship to engage public interest. But suggestive or lewd reporting is a problem: for scientists, because their work is misrepresented; for sexual minority groups, because it equates their existence to an illness; and for society, because the articles feed negative stereotypes. In a world in which discrimination against sexual minority groups remains common, scientists should present their work objectively, and do all they can to avoid its misappropriation and misuse.

To analyse how the media reports on sexual behaviour in animals other than humans, we surveyed¹ 48 newspaper, magazine and Internet articles written about 11 papers (see 'Science and sensationalism' for examples). We excluded blogs that presented the opinion of the blog author, and any article published on sites with obvious pro- or anti-gay, lesbian or transgender agendas. We found consistent patterns in how sexual-behaviour research is misrepresented and propose suggestions for how it might be avoided.

PAIR BONDS

The vast majority of studies reporting sexual contact between pairs of males or females were presented in media articles as documenting gay, lesbian or transgender behaviour. This is not innocuous — these are terms that refer to human sexuality, which encompasses lifestyle choices, partner preferences and culture, among other factors.

More worryingly, studies that invoked atypical sexual behaviour through genetic or hormonal manipulation were reported as inducing gay or lesbian behaviour or changing the animals' sexual orientation, even in the case of the nematode *Caenorhabditis elegans*, which has males and hermaphrodites, rather than males and females. ►

► To the general public, such inaccurate coverage implies that homosexuality is some sort of illness, which marginalizes a section of human society.

A study of the neurobiological features associated with male–male sexual behaviour in domestic rams² was reported as an effort to “cure” homosexuality in sheep, which “could pave the way for breeding out homosexuality in humans”³. This reporting led to extensive and coordinated protests against the researchers and their institution from both gay-rights and animal-rights activists, including an e-mail campaign with more than 20,000 signatories calling for an immediate halt to the research programme⁴.

Concerns about responsible media coverage have been raised with respect to studies in humans of race and IQ. In research into sexual behaviour in animals other than humans, consistently polarized reporting is the norm.

CHOOSE WORDS CAREFULLY

What can be done to promote more effective and constructive communication of this topic? Careful, objective scientific writing isn't enough. The sensationalist titles of the media articles had no clear connection with the tone of the original papers. Even papers

that had restrained and accurate titles were outrageously misinterpreted. For example, ‘Female-limited polymorphism in the copulatory organ of a traumatically inseminating insect’⁵ became ‘Bat bugs turn transsexual to avoid stabbing penises’⁶.

The direct quotes that researchers give to journalists seem to have a large influence on the tone of the articles written. Interviews in which scientists drew a link between their research findings and human behaviour consistently led to more-inflammatory media articles. By contrast, scientists who gave no such indications seemed to avoid the worst sensationalism.

These findings suggest that scientists can shape the coverage of their results. Perhaps the most striking examples of successful communication were the articles that came from research led by Lindsay Young, a wildlife biologist with Pacific Rim Conservation in Honolulu, on the breeding behaviour of the Laysan albatross (*Phoebastria immutabilis*)⁷. Young was regularly quoted as saying “Lesbian is a human term. The study is about albatross. The study is not about humans”. When asked what her study said about human behaviour, Young's only quoted reply has been “I don't answer that question”⁸.

As a result, most of the media coverage of this research used the term ‘same-sex couples’ when referring to albatross pairing, and only one used ‘lesbian’. More significantly, most of the coverage spent more time discussing the behaviour of the birds than speculating on its relevance to humans or poking fun at the findings. In this case, actively denying inappropriate speculation seems to have helped to restrain the tone of the articles without diminishing public interest in the work.

Research on sexual behaviour in animals does not need to be sensationalized to catch public attention. Scientists should be encouraging public interest in their work, and engaging in the debates that surround social and cultural attitudes to human sexuality. But they must also try to ensure responsible reporting of their findings through clear, carefully phrased messages to journalists. The most important advice to scientists working in any area is to maintain a consistent and objective line throughout the paper and in all interactions with the press.

Ultimately, any one study can tell us about the sexual behaviour of only the species under investigation. It might also provide an avenue for exploration in humans, and for phylogenetic analyses of the evolution of sexual behaviour. But simplistic extrapolation, by scientists or by the media, to ‘explanations for’ human heterosexual, gay, lesbian or transgender behaviour can only stand in the way of these worthy and exciting goals. ■

Andrew B. Barron is a senior lecturer at the Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia. **Mark J. F. Brown** is a Reader in evolutionary ecology and conservation at the School of Biological Sciences, Royal Holloway University of London, UK. e-mail: andrew.barron@mq.edu.au

1. Bailey, N. W. & Zuk, M. *Trends Ecol. Evol.* **24**, 439–446 (2009).
2. Perkins, A. & Roselli, C. E. *Horm. Behav.* **52**, 70–77 (2007).
3. Oakeshott, I. & Gourlay, C. *The Sunday Times* (31 December 2006).
4. Schwartz, J. *New York Times* (25 January 2007).
5. Reinhardt, K. et al. *Am. Nat.* **170**, 931–935 (2007).
6. Hooper, R. *New Sci.* (19 September 2007).
7. Young, L. C. et al. *Biol. Lett.* **4**, 323–325 (2008).
8. Mooallem, J. *The New York Times* (31 March 2010).
9. Shine, R. et al. *Nature* **414**, 267 (2001).
10. Milius, S. *ScienceNews* **160**, 311 (2001).
11. Saletan, W. *The Washington Post* (4 February 2007).
12. Fox, M. *Reuters* (4 November 2005).
13. Cloud, J. *Time Magazine* (26 January 2007).
14. White, J. Q. et al. *Curr. Biol.* **17**, 1847–1857 (2007).
15. Casselman, A. *Natl Geogr.* (25 October 2007).
16. Dolgin, E. *Nature* dx.doi.org/10.1038/news070917-7 (2007).
17. Levan, K. E. et al. *J. Evol. Biol.* **22**, 60–70 (2008).
18. Holmes, B. *New Sci.* (20 October 2008).
19. Brennan, Z. *Daily Mail* (18 May 2010).
20. Mooallem, J. *Independent* (1 May 2010).

SCIENCE AND SENSATIONALISM

Six examples of how journal articles dealing with research on sexual behaviour in animals have been covered in the media.

Journal article	Media headline	Phrases used in media article
Advantage of female mimicry to snakes ⁹	She-male Garter Snakes: Some Like It Hot ¹⁰	“Male garter snakes can mimic females and drive dozens of other guys to wriggle over them. The force behind this deluded orgy may not be sex, though.”
The ram as a model for behavioral neuroendocrinology ²	Brokeback Mutton ¹¹ Gay sheep may help explain biology of homosexuals ¹² Yep, They're Gay ¹³	“Gay rams don't act girly. They're just as gay in the wild.” “Gay sheep that mate only with other rams have different brain structures from ‘straight’ sheep, a finding that may shed light on human sexuality.” “Another small but fascinating finding: all gay rams are butch.”
The sensory circuitry for sexual attraction in <i>C. elegans</i> males ¹⁴	Sexual Orientation Is Genetic In Worms, Study Says ¹⁵	“The sexual preferences of nematode worms can be changed by flipping a genetic master switch in their brains.”
Female-limited polymorphism in the copulatory organ of a traumatically inseminating insect ⁵	Bug sexual warfare drives gender bender ¹⁶	“Females are retaliating by imitating males.” “For these females, perhaps it's a stark choice between dressing as a male, or dressing their wounds.”
Testing multiple hypotheses for the maintenance of male homosexual copulatory behaviour in flour beetles ¹⁷	How gay sex can produce offspring ¹⁸	“Homosexual activity might, for example, help males practise for straight sex.”
Successful same-sex pairing in Laysan albatross ⁷	Lesbian albatrosses, gay giraffes and some very open-minded penguins. So, can animals really be gay? ¹⁹ The love that daren't squawk its name: when animals come out of the closet ²⁰	“Many of the albatrosses appear to be, well ... gay.” “Young would never use the phrase ‘straight couples’. And she is adamantly against calling the other birds ‘lesbians’ too.”



A cartoon from 1892 proposes America close its ports to prevent the influx of disease from abroad.

PANDEMICS

A deadly business

Andy Tatem traces the global tracks of pathogens that have clung to the coat-tails of trade over the centuries.

In 1866, the Archbishop of Canterbury called for Britain to observe a national 'day of humiliation' to atone for the sins that had triggered recurring epidemics of disease. (Such days now occur only when British football teams exit international tournaments yet again.) It wasn't, of course, divine retribution that had brought plague, yellow fever, cholera and rinderpest to the United Kingdom. Rather, it was the expansion of trade networks.

Disease has always followed the world's main arteries of trade. Plague, for instance, was spread by merchant traders in Europe throughout the seventeenth century. Huge yellow-fever outbreaks in the Americas in the nineteenth century were triggered by the slave trade. More recently, outbreaks of respiratory diseases, including avian influenza, have escalated rapidly into pandemics as a result of air travel. In *Contagion*, historian Mark Harrison chronicles the deadly legacy of commerce, and humanity's efforts to cope.

Harrison recounts how nations have attempted to protect themselves from disease, while recognizing that sanitary measures can hurt trade. After city officials admitted a plague outbreak in Oporto, Portugal, in 1899, for instance, it was noted that "it is no light matter to proclaim the existence

of plague in a sea-port town dependent for its prosperity on maritime commerce".

And before modern regulations, ethics could be at odds with the need to make a profit. When, if at all, should a disease outbreak be announced, given the economic consequences of quarantines by trading partners? How can officials convince trading partners that an outbreak is being tackled effectively? How should a nation react to outbreaks in nations they trade with? *Contagion* is full of colourful examples of the various responses to these questions from the past 700 years.

Heavy-handed quarantine was often used to damage rivals' economic prospects. Frosty relations between Britain and Egypt in the late nineteenth century were exacerbated, for instance, when Egypt imposed harsh quarantine measures on ships sailing through the Suez Canal to Europe from cholera-endemic regions such as India. Britain's foreign office stated that these measures "serve no better



Contagion: How Commerce Has Spread Disease
MARK HARRISON
Yale Univ. Press: 2012.
416 pp. £25, \$38.00

purpose than the enriching of a set of incompetent and corrupt officials". But Britain's behaviour had at times been no better: in the eighteenth century, the government imposed 40-day quarantines on all ships from the Mediterranean, and fumigated goods with damaging chemicals. This 'plague control' probably had more to do with gaining economic advantages than with halting the spread of disease, and resulted in revolts and conflict.

Claims of protectionism are common in *Contagion*. When diseases arise, importing nations often impose sanctions, from plague-related tit-for-tat trade embargoes by rivals Britain and the Netherlands in the seventeenth century to European bans on US pork in the past decade. An outbreak of fire blight in New Zealand apples in the 1920s prompted Australia to enact a ban on the fruit that lasted for almost 90 years. In 2005, Jim Sutton, then New Zealand's agriculture minister, raged that "Australians cheat in matters of biosecurity and the concept of honest science has no meaning to them".

Among the many histories of pandemics, *Contagion* stands apart for its depth of research and breadth of sources, and its focus on commerce. The narrative is well balanced between the high-level political decision-making surrounding sanitary measures and more focused examples of how individual cities, traders or ships were affected.

If this excellent book has a weakness, it is that it often overlooks the basic epidemiology of trade-borne diseases, leaving most readers guessing as to whether the debated interventions were effective.

As shown by *Contagion*, in an age of global connectivity, governments must tread a fine line between protecting public health and ensuring economic prosperity. This can create delicate situations. With the continued expansion of air travel and economic globalization, historical lessons are hugely relevant. Profit-seeking and public health can still be at odds, and political and commercial interests often take priority over scientific decisions. As they have been for centuries, nations are torn between offending their trading partners and alarming their citizens.

The dilemma is illustrated by the case of violent protests that broke out in South Korea in 2008, costing the country around US\$2.5 billion, when US beef imports were resumed after a 5 year ban following a case of bovine spongiform encephalopathy (BSE). As a 13-year-old South Korean schoolboy put it, "I am afraid of American beef. I could study hard in school. I could get a good job and I could eat beef and just die." ■

Andy Tatem is an associate professor at the University of Florida in Gainesville. He studies the geography and dynamics of human and pathogen movements.
e-mail: andy.tatem@gmail.com



Humans are surprisingly adaptable: Philippa Smeed, who was born without arms, supports a mug with her foot.

TECHNOLOGY

Beyond the body

Ewen Callaway finds immersion in human enhancement to be both unsettling and uplifting.

S*uperhuman*, a provoking and discomfiting exhibition at the Wellcome Collection in London, projects equal parts Clark Kent and Superman. An exploration of human enhancement, it starts with early-nineteenth-century tortoiseshell spectacles and an ancient bronze Icarus to remind us that the concept is neither inherently high-tech nor particularly new.

Curator Emily Sargent takes a broad view of technological improvement. Vivienne Westwood platform heels, Viagra pills and news clippings of the first child born as a result of *in vitro* fertilization are displayed beside cochlear implants and an ivory dildo. Sex is everywhere in *Superhuman*.

The exhibition explores the tensions between replacement and improvement, form and function. Should technological aids — be they pills or prostheses — restore ‘normality’ (whatever that means) or transcend it, like Icarus’ wings? Is fidelity more important than utility?

The ancients favoured form. A perfectly sculpted big toe made of a composite of linen and gesso was found alongside an Egyptian mummy dating to the sixth century BC. Historians believed at first that it was meant only as a replacement in the afterlife, but the prosthesis fit snugly on the feet of other people missing their big toes, suggesting that it may have found a use in the real world, too. Similarly, a silver prosthetic nose welded to spectacle frames is testament to attempts by a well-heeled woman of the Victorian era to avoid social stigma after losing her nose to syphilis.

Yet too much emphasis on assimilation can have heartbreaking consequences. Between 1957 and 1962, many pregnant women took the drug thalidomide to ease morning sickness and improve sleep, until doctors realized that it caused severe limb deformities in the developing fetus. The UK government funded the design of prosthetic arms and legs for children born without them. The disturbingly lifelike limbs were so

heavy and clunky that some had to be powered by pneumatic pumps, and they were mostly useless. In one

of a series of haunting videos in the exhibition, a child describes using the prostheses as a weirdly dreamlike experience: “It was like somebody else touching [things] and I was merely an observer.”

Beside the display of videos is a black-and-white photograph of an 11-year-old girl, Philippa Smeed (now Verry), who was born without arms, lounging on a sofa and sipping from a mug that dangles from her toe. Her eyebrows are arched, and the mug seems to conceal an impish grin. “When I look at this photograph I realise what a wonderful, confident and happy childhood I had,” Verry says.

Examples of the extravagant and strange abound in *Superhuman*. There are cyborgs with limbs and faces cobbled together from workshop tools. There is athlete and double-amputee Aimee Mullins, made up as a cheetah with prosthetic hind legs: a still from artist Matthew Barney’s film *Cremaster 3*, which will screen at the exhibition. And there is the Whizzinator, a combination penis and bladder prosthetic, designed to help male athletes to evade anti-doping tests.

Aptly, *Superhuman* opened days before the start of the London Olympics. Sport is a study of human enhancement, by means

Superhuman
Wellcome Collection,
London.
Until 16 October
2012.

both honourable and underhand. South African sprinter Oscar Pistorius, who was born with missing leg bones, is running in both the Olympics and the Paralympics. His carbon-fibre Cheetah prosthetic blades, fitted with running spikes, look nothing like human limbs, but they perform so well that the International Association of Athletics Federations appealed (unsuccessfully) against Pistorius's participation in races against able-bodied athletes, claiming that he had an unfair advantage.

At the start of the twentieth century, Olympic marathon runners were allowed to train intensively for only four weeks in the year before the Games. They used strychnine as an endurance enhancer to help them to complete the gruelling course of about 42 kilometres (26 miles). A blurry photograph, on display in *Superhuman*, shows a dazed Tom Hicks, winner of the marathon in the 1904 games in St Louis, Missouri, being helped to the finish line by his two trainers.

Like other exhibitions at the Wellcome Collection, *Superhuman* leans heavily on historical and contemporary objects, and primary documents such as newspaper clippings. But the few fine-art displays are high points. South Korean artist Hyungkoo Lee's photograph *Enlarging my Right Hand with Gauntlet* depicts a pump resembling an udder.

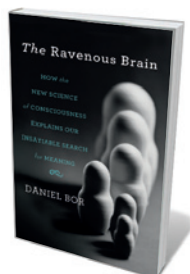
In her tongue-in-cheek short film *I Need a Hero*, Charlotte Jarvis imagines a reality-television show in which amputees compete for body modifications and enhancements. The most powerful film is Regina José Galindo's *Recortepor la Línea* (Cut Through the Line). In it, the artist stands naked and taciturn in a public park in Venezuela — home to the world's third-highest rate of cosmetic surgery per capita — as a plastic surgeon uses a marker to chart the changes he would make to her body.

Revital Cohen's room-sized installation *The Immortal* is the most ambitious artwork: a closed loop in which air and water flow between a respirator, an incubator, a dialysis unit and other medical machinery, in an attempt to simulate a living organism. It will run for just one hour a day, so plan your visit accordingly.

The exhibition ends with a series of video debates on topics such as cognition-enhancing drugs, life-span extension and transhumanism — the idea that humans can transcend their bodies through technology. My favourite is a meditation by bioethicist John Harris on the obligation to enhance. Humans, he argues, must get over their squeamishness if they are to outlive climate change, global pandemics and eventually the destruction of Earth itself. Sounds like a job for Superman. ■

Ewen Callaway is a news reporter for *Nature* in London.

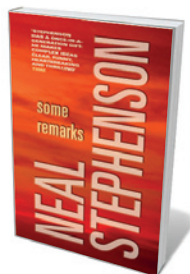
Books in brief



The Ravenous Brain: How the New Science of Consciousness Explains Our Insatiable Search for Meaning

Daniel Bor BASIC BOOKS 352 pp. £18.99 (2012)

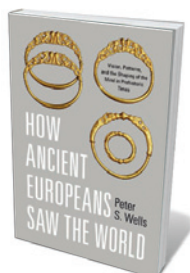
As scientific enterprises go, cracking consciousness is up there with deciphering dark matter. Neuroscientist Daniel Bor dives into the conundrum with relish. He begins by defining consciousness as the ability to gather knowledge, then works his way from a history of the brain and the “neuroscience of awareness” to an exploration of severe brain damage. Intriguing arguments abound — not least, Bor's recasting of mental conditions such as schizophrenia and bipolar disorder as ‘disorders of consciousness’.



Some Remarks: Essays and Other Writing

Neal Stephenson ATLANTIC 336 pp. £20 (2012)

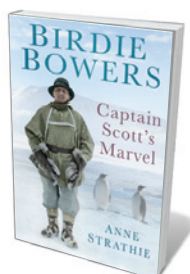
From *Snow Crash* (Bantam, 1992) to *Reamde* (Atlantic, 2011), Neal Stephenson's novels range over a dazzling array of disciplines — including metaphysics, gaming, nanotechnology and the history of science. Here, he assembles an entertaining sampler of cyberpunkish treats. Among freshly edited essays, interviews and other short works on topics such as geek cool and the mainstreaming of science fiction are two previously unpublished pieces. ‘Get Up’ is an essay on sitting; the other is a work of fiction one sentence long. Prepare to be amused.



How Ancient Europeans Saw the World: Vision, Patterns, and the Shaping of the Mind in Prehistoric Times

Peter S. Wells PRINCETON UNIVERSITY PRESS 304 pp. £24.95 (2012)

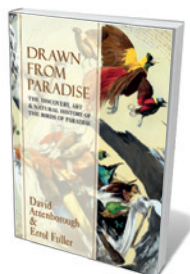
Look at a pot predating the Roman Empire, or a Bronze Age burial site, and your interpretation of form and pattern will be vastly at variance with that of their makers. So says anthropologist Peter Wells, who argues that in late prehistoric Europe — a world that lacked the written word and was thin on ‘stuff’ — people's perceptions were very different from our own. Wells ‘reads’ tools, vehicles, ornaments, textiles and buildings to reveal a neurobiological map of profound changes in ancient society.



Birdie Bowers: Captain Scott's Marvel

Anne Strathie THE HISTORY PRESS 224 pp. £18.99 (2012)

Scott of the Antarctic was surrounded by strong characters. One, Henry ‘Birdie’ Bowers, emerges as grounded, courageous and ferociously well organized. He handled landing, navigation and more for the 1910–12 Terra Nova expedition, and took part in the penguin-egg quest immortalized in Apsley Cherry-Garrard's memoir *The Worst Journey in the World* (1922). He was also the youngest to die with Scott in the doomed ‘party of five’. Built on research in Antarctica and at the Scott Polar Research Institute in Cambridge, UK, Anne Strathie's biography includes previously unpublished material.



Drawn From Paradise: The Discovery, Art and Natural History of the Birds of Paradise

David Attenborough and Errol Fuller COLLINS 256 pp. £30 (2012)

Some dance; some are dubbed superb or magnificent; many sport ruffs, streamers or elaborate headgear. Birds of paradise — found in New Guinea and Australia and comprising about 40 species — are celebrated here by broadcaster David Attenborough and artist Errol Fuller. The authors trace the natural history of these beauties, known to Europeans since the sixteenth century, with copious illustrations by artists from Peter Paul Rubens to Jacques Barraband.

Correspondence

To embrace doping in sport is absurd

I find your discussion of performance-enhancing drugs (*Nature* **487**, 287–289; 2012) disrespectful to the millions of elite and sub-elite athletes who rely solely on training to reach their goals. In reality, most of these drugs do not create “superhuman athletes” but allow athletes to take short cuts in training or to recover faster from injury.

The sport of running has been plagued by doping for years, and a slew of positive tests has been reported in recent weeks. But to suggest that doping should be embraced because drug-testing efforts are a losing battle is absurd. For every performance-enhancing-drug pragmatist, there are many who would rather see stricter testing regimens and harsher penalties for offenders.

Science does hold promise for future athletes wanting to break today's barriers, but the realms of exercise science, sports psychology and nutrition are likely to have more bearing than pharmaceuticals and genetics (*Nature* **487**, 297; 2012).

Ryan Purcell Emory University, Atlanta, Georgia, USA.
ryan.purcell@emory.edu

Misconduct rule is not retroactive

Octavian Voiculescu is concerned that the latest regulations of the Romanian National Ethics Council might be used retroactively, which would be unconstitutional (*Nature* **486**, 34; 2012). As president of Romania's National Authority for Scientific Research that helped to introduce these regulations in 2011, I would like to stress that this is not the case: the regulations are not implemented retroactively.

The regulatory changes include a provision to terminate automatically the employment contracts of academics or

researchers found guilty of “grave misconduct” — plagiarism, falsification of scientific data or false statements in grant applications.

The non-retroactive legal principle means that no misconduct committed before the legislation came into effect can be penalized on the basis of that legislation. This principle has been strictly observed under my administration.

It is incorrect to interpret the non-retroactive principle as meaning that sanctions cannot remove rights or privileges that were in operation before the legislation came into effect, with people losing their jobs only if they were appointed after 2011.

It is not the date of employment that matters, but the date on which the illegal action was undertaken: if the action was perpetrated before 2011, it cannot be sanctioned; if it happened afterwards, it can.

Dragoş Ciuparu Petroleum-Gas University of Ploieşti, Romania.
dciuparu@upg-ploiesti.ro

Put patients and researchers in touch

As Research Network leader at the UK Alzheimer's Society, I can vouch that the value of the biomedical research community engaging with patients and carers acts both ways (*Nature* **487**, 7; 2012).

Research supported by the Public and Patient Involvement programme of the UK Alzheimer's Society couples scientific and clinical expertise with the insight of people who are affected directly or indirectly by dementia (www.alzheimers.org.uk/researchnetwork). These people convey their research priorities to us; we offer them an equal presence and vote at funding panels.

For every grant we commission, we assign three carers, former carers or people with dementia to provide an informal monitoring group who

visit the researchers once or twice a year. These volunteers encourage the research team by sharing their own experiences, and help us to improve our protocol, patient communication and recruitment methods.

The patients and carers learn more about the complexity, challenges and timescales of biomedical research. The researchers come to realize that dementia is about much more than memory loss, and that sufferers are still human beings.

This two-way communication motivates patients and carers to fund-raise and campaign for the Alzheimer's Society. Since 1999, we have awarded more than £15 million (US\$24 million) to these research partnerships.

Matt Murray Alzheimer's Society, London, UK.
matt.murray@alzheimers.org.uk

Alarming shift away from sharing results

There have been worrying attempts in the past year to tamper with the sharing of scientific research results. Each seems indicative of a shift away from the classical principles of science.

There was the review of two H5N1 avian influenza virus studies in ferrets by the US National Science Advisory Board for Biosecurity in December 2011. The board initially recommended that the published studies should “not include the methodological and other details that could enable replication of the experiments by those who would seek to do harm” (see go.nature.com/nywkdy). Fortunately, this attempted ban was later withdrawn.

In two disturbing examples of publication irregularities, a meta-review of biomedical papers in leading journals revealed that key primary data are not always made publicly available (A. A. Alsheik-Ali *et al.* *PLoS ONE* **6**, e24357; 2011); and

Bernardo Huberman alerted the community to the practice of using non-disclosed data from private sources (*Nature* **482**, 308; 2012).

In December 2011 a bill was proposed to the US Congress to reverse the National Institutes of Health policy that all taxpayer-funded research should be freely accessible online (see go.nature.com/uvj68l). The bill's proponents later withdrew their support, but legislative action would have severely limited the diffusion of scientific knowledge.

Science should be available for evaluation by other scientists and for public scrutiny, just as it has been since Galileo's time. It should not be heading for epistemological suicide as a result of vested interests or a creeping loss of awareness of the theory of knowledge.

Giovanni Boniolo University of Milan and European Institute of Oncology, Milan, Italy.

Thomas Vaccari FIRCI Institute of Molecular Oncology (IFOM), Milan, Italy.
thomas.vaccari@ifom.eu

Time for a Higgs metaphor upgrade

I liked your sidebar ‘What is the Higgs?’ (*Nature* **487**, 147–148; 2012). This offers the first hint that there may be new explanations beyond the standard metaphor that it “endows particles with mass”. After reading words to that effect in so many news articles that simply leave it at that, this scientifically literate layman is ready for a metaphor upgrade.

If this turns out to be ‘the end’ for theoretical physicists, maybe they could find gainful employment working with science writers to develop metaphors that match the levels of scientific and mathematical sophistication of a wider range of target audiences.

Charles Packer Washington DC, USA. mailbox@cpacker.org

WAYS TO ENJOY NUTRIENT BLEND 14

A taste of real life.

BY LUC REID

A set of marble stairs led up into the apartment building through tall bronze doors. Inside, shifting sunlight filled an anteroom, carrying colours down from stained-glass skylights. In the distance John heard music — lutes, maybe, and pipes. The air was perfumed with new leaves, water, lilacs.

John climbed the steps, reluctant for no reason he could name. Finally he'd wheedled an invitation to one of the Garrets' dinners, a parade of wonders. Unless he was suddenly tired of marvels, he should be excited.

Other guests were also arriving. He forced a smile to avoid bothering them.

A slave 20 feet away bowed and called out to him: "Citizen Kindersley! My masters welcome you!"

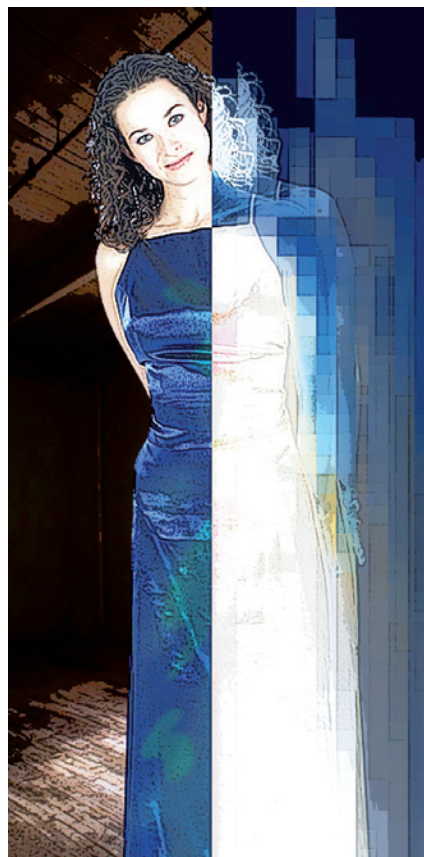
The slave was too much for John and, in a completely irrational fit of annoyance, he switched off his augmented-reality system, something he hadn't done in he didn't remember how long.

The marble, the stained glass, the slave and the statues evaporated to reveal close, badly painted cinderblock walls lit by a flickering fluorescent lighting fixture decades older than he was. Scents of leaves and lilacs gave way to odours of dust and mildew. The slave wasn't a real person, unsurprisingly, and even the space he occupied was false: the real room was much smaller.

He was about to switch augmentation back on when he was distracted by a glimpse of something that wasn't grey or brown. One of the other guests, a calm-looking woman with unruly brown hair, was wearing a bright blue dress. The colour was a little uneven, suggesting she'd dyed it herself, but it was striking anyway, given the dull, faded and stained look of everyone else's clothes, including his. He looked down at himself and powered augmentation back up.

Back came the marble, the colours, the scents and the slave, who said: "This way please, sir," and indicated the stairs. He was blocking John's view of the woman in the blue dress, and John leaned around him to get a glimpse. She was just disappearing up the steps. In virtual, her dress was unremarkable: white, vaguely Roman-looking — and clearly at least four or five weeks out of date. Didn't she care what she looked like? Who didn't have a few minutes to

GO NATURE.COM
Follow Futures on
Facebook at:
[go.nature.com/mtoodm](https://www.facebook.com/mtoodm)



update fashions every morning? Yet somehow he envied her. He was envying neglect or social ineptitude, he told himself, but the feeling didn't go away.

Her graceful walk and bright dress played in his head in sharp detail as he climbed the stairs.

The Garrets were greeting the arrivals at the top, laughing as they rode from guest to guest on gold cushions held up by straining cherubs. Their 'apartment' had an expansive mosaic floor and marble pillars under a vaulted roof, the whole arrangement open to the air on three sides. John looked out to see a verdant tumble of hills patterned with woods, farms, streams and flower gardens. Around the guests capered jugglers and fire-eaters, while musicians plucked and winded their instruments, and the table in the centre of the room strained under an eye-popping feast. John distractedly took a seat, searching for the woman with the hidden blue dress, but waiters and performers made it difficult to see clearly. Finally he switched augmentation back off, showing him to be in a crowded one-room apartment

with three rickety card tables surrounded by mismatched kitchen chairs and set with large platters of what looked like Nutrient Blend 14. John hadn't had Blend 14 without augmentation for a long time, but he remembered it being bland and just a little bitter, with a texture like soggy asparagus.

There she was, at the far end of the room, sitting quietly. Around her, guests gaped into the air at the Garrets' marvels. She herself was not gaping: she was looking directly at him.

John started as the other guests burst out laughing and applauded, staring at a nearby dish. He switched augmentation back on and caught sight of the last few birds flying out of a steaming pie. The pie then miraculously filled up with chopped meat, nuts and spices, and Vlad Garret began serving pieces out at random, including to John. John took a bite and grunted with appreciation: the crust was buttery and flaky, the meat rich and succulent, the spices harmonious and complex. Taken all together, it was nearly indecent.

The waiters parted for a moment, and he saw that the woman was no longer sitting where she'd been a moment before. He switched off augmentation, transmuting his pie to bitter mush in his mouth. He swallowed. Someone tapped him on the shoulder, and he turned. It was her.

"Enjoying yourself?" she asked.

"More now," he said. It was a ridiculous statement, considering, but it was true. He studied her, wondering what it was about her that was strange. Then he had it: no augmentation wires going into her temples. She had a small eViewer dangling around her neck, but she wasn't wired, although he could see the ports where she must have been connected once.

What was it like, spending your days seeing people reacting to things that weren't there, wandering through a dingy grey cityscape, the food bland, the scents wretched? Why had she chosen *that*?

She held out her hand, and John took it and stood up. It was warm, and her fingers curled around his as he took them. He'd been thinking the choice she'd made had been to live in the real world instead of the beautiful one. Now he was thinking that might not be what the options were at all. ■

Luc Reid writes fiction and nonfiction and speaks on the psychology of habits. His articles, books and short fiction can be found at lucreid.com.

FORUM: Astrobiology

Frontier or fiction

Astrobiology, the study of life in the Universe, is sometimes criticized as being a fashionable label with which to rebrand existing research fields. Its practitioners, however, argue that the discipline provides a broad framework for developing a better understanding of the frontiers of biology. A biologist and a planetary scientist offer their views.

THE TOPIC IN BRIEF

- Space-exploration programmes in the 1960s powered a new branch of biology called exobiology, which was focused on the search for life beyond Earth.
- Today, the term exobiology has been almost abandoned, to be replaced by a broader one, astrobiology — the study of the origin, evolution, distribution and future of life in the Universe¹.

- However, the nature of astrobiology as a distinct scientific discipline has been challenged because of the lack of proof for extraterrestrial life.
- Also, astrobiology is sometimes criticized as being a buzzword added to certain research topics in fields such as planetary science or biology — for example, in the study of microbes known as extremophiles, which live under extreme conditions.

Maintaining the plausible

ANTONIO LAZCANO

Stimulated by controversies² that arose over the evidence for extraterrestrial life in the Martian meteorite ALH 81004 in the mid-1990s, NASA reorganized its programmes on exobiology and planetary science as part of an attempt to integrate its research on life sciences into its space-exploration efforts. Thus was born astrobiology, an all-encompassing effort that was expanded to include the study of extremophiles, the occurrence of planets and their habitability beyond our Solar System, and research on the origin and evolution of life on Earth, among other disciplines³. But can a funding programme be transformed into a new science?

The astrobiology programme has indeed supported such studies, and has created jobs for young researchers, endorsed the teaching of evolutionary biology, and promoted major efforts in outreach following NASA's generous tradition of sharing its scientific results for free. The creation of the NASA Astrobiology Institute¹ was followed by the launch of specialized journals and scientific societies, university courses, graduate programmes and books, as well as by a handful of centres and networks of varying scope and uneven academic standards in countries other than the United States. The commitment of such efforts to understanding evolutionary perspectives is

a major accomplishment, but I feel that many of them tend to place far too much weight on a handful of loose analogies between extremophilic microbes and the potential habitability of other worlds in our Solar System.

In the absence of unambiguous proof for its existence, almost nothing can be said about extraterrestrial life about which the opposite is not also true. The scarcity of evidence gives considerable latitude, and, in certain circles, astrobiology has become a resounding but meaningless catchword in the competition for grant money. It has been argued that the poten-

"In certain circles, astrobiology has become a resounding but meaningless catchword in the competition for grant money."

tial discovery of a terrestrial 'shadow biosphere' — that is, organisms that have an alternative chemical composition to that of all known organisms — would imply that life has appeared more than once on our planet and therefore that it could also have developed in other worlds.

In spite of the fanfare⁴ that advertised the existence of 'arsenic-based DNA' in a microbe isolated from a saline lake in California⁵, mere distilled water and additional laboratory controls have washed away the speculation about such alien biochemistries^{6,7}. And some of the attempts by other researchers to extrapolate to other parts of the Universe the ability of microbes to adapt to extreme environments may be due more to the struggle for funding

than to the desire to study habitable planets.

The search for life beyond Earth is a legitimate scientific question and an alluring intellectual endeavour that can best be served by keeping a healthy distance from science-fiction scenarios and from the theological musings that somewhat surprisingly find their way into astrobiology meetings. Depending on who you speak to, astrobiology seems to include everything from the chemical composition of the interstellar medium to the origin and evolution of intelligence, society and technology — as if the Universe is following an inevitable upward linear path leading from the Big Bang to the appearance of life and civilizations capable of communication.

Neither the formation of planets nor the origin of life is seen today as the result of inscrutable events; rather, they are considered as natural outcomes of evolutionary processes. However, this does not mean that such outcomes are inevitable, and it is still to be shown that life exists — or has existed — in places other than Earth (Fig. 1). The lack of evidence should not inhibit us in the slightest, but unless we are bound by the highest academic standards and critical attitudes, astrobiological discussions will become nothing more than empty speculation laced with a formidable disregard for scientific plausibility.

*Antonio Lazcano is a biologist in the Facultad de Ciencias, Universidad Nacional Autónoma de México, Ciudad Universitaria, 04510 Mexico DF, Mexico.
e-mail: alar@ciencias.unam.mx*

The last great experiment

KEVIN P. HAND

Of the four basic sciences — physics, chemistry, geology and biology — only one has yet to prove its functionality beyond Earth. In the centuries since Galileo, we have come to learn that the laws of physics, and the

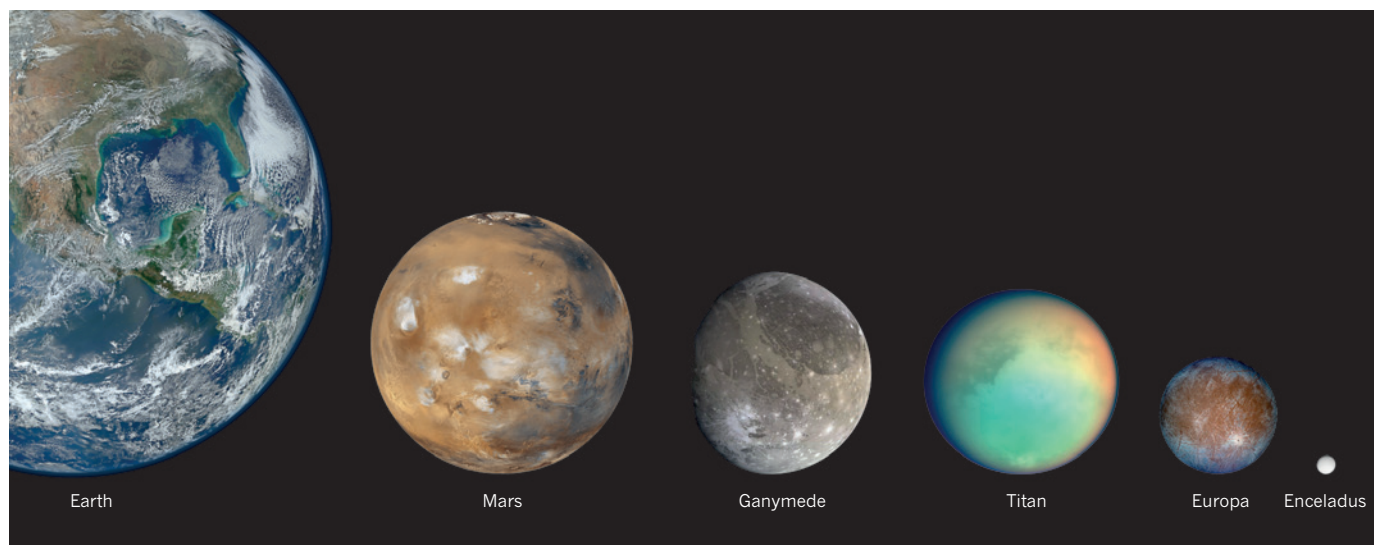


Figure 1 | Life around us? Searching for extraterrestrial life is one of the aims of astrobiology, and there are several potentially habitable worlds in the Solar System, in addition to Earth. Two of Jupiter's moons (Ganymede and Europa) and two of Saturn's moons (Titan and Enceladus) harbour liquid-water oceans. And Mars probably had vast liquid-water environments during much of its early history.

principles of chemistry and of geology, work beyond our planet. Yet when it comes to this bizarre phenomenon known as life, we have yet to make that leap. The interdisciplinary field of astrobiology seeks to address that goal through the exploration of other worlds and of the limits of life on Earth.

A common misconception is that astrobiology is equivalent to the search for life elsewhere. Some people have even gone so far as to say that it is a science without a subject because we don't yet have any evidence for extraterrestrial life^{8,9}. But that is a flawed argument that has been put to rest on several occasions (see, for example, ref. 10). Many experiments in science target hypothetical particles or objects; biology is simply handicapped by the fact that first principles and mathematics provide limited predictive power. In fact, the search for life beyond Earth is just one subset of astrobiology. As Knoll *et al.* have written¹¹: "astrobiology can be thought of as the application of geobiological principles to the study of planets and moons beyond the Earth."

The heavily debated detection of methane on Mars serves as a useful example¹². Because methane is rapidly destroyed by ultraviolet light from the Sun, the alleged presence of this molecule in Mars's atmosphere would necessitate an active source from geological or biological processes. Astronomers, chemists, geologists and biologists have all weighed in to address the issue from different perspectives and have worked together in an effort to triangulate on the major strengths and weaknesses of the data and models. Considerable support for that interdisciplinary work comes from NASA's Astrobiology Institute. This summer, the Mars Science Laboratory (MSL; a new NASA mission to explore the red planet) may help to resolve this debate, and the astrobiology community will be able to put the MSL

results into context in part owing to our understanding of the geochemical and biological processes that generate methane here on Earth.

Some have argued that we know so little about life on Earth and its origin that we should focus our limited resources on these challenges⁹. Sixty years ago I might have been mildly sympathetic to this viewpoint,

"If funding permits, within the next few decades we will know the answer to whether or not life exists elsewhere in our Solar System."

but now I think it's fair to say that our mapping of the tree of life on Earth begs the question of whether other trees exist. The analogy in chemistry would be as if, after creating the first periodic table in the 1860s, Mendeleev and others had decided not to search for more elements, even though the gaps in the table provided a guide for where to look. Life on Earth serves as a guide for identifying potentially habitable environments elsewhere. Yes, there is still much to learn about the specifics of terrestrial organisms, but our understanding of life as a phenomenon and of biology as a science will be greatly advanced by finding a second, separate origin that can help to put what we observe here on Earth in context.

And we now know where to go to conduct this great experiment. The vast, global subsurface liquid-water ocean of Jupiter's moon Europa¹³ is arguably the best place to search for extant life in the Solar System. Furthermore, Europa's great distance from Earth nearly ensures that the two worlds have not seeded each other. Any life on Europa would thus represent a second, independent origin, even if it happened to converge on the chemistry based

on DNA, RNA and proteins found in Earth's organisms.

If funding permits, within the next few decades we will know the answer to whether or not life exists elsewhere in our Solar System. We are perhaps just two or three missions away from having received enough data back from several potentially habitable worlds (Fig. 1) to know whether or not life has ever taken hold as a widespread process on other planets or moons. The combined cost of such missions would be comparable to that of the Large Hadron Collider, which has just revealed evidence for the Higgs boson, completing one of the other last great experiments of our era. It's time for biology's great experiment. It's time to learn whether we live in a biological Universe or one in which life on Earth is a singularity. ■

Kevin P. Hand is a planetary scientist at the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91103, USA. e-mail: kevin.p.hand@jpl.nasa.gov

1. <http://astrobiology.nasa.gov/nai>
2. Darling, D. *Life Everywhere: The Maverick Science of Astrobiology* (Basic, 2001).
3. Dick, S. J. & Strick, J. E. *The Living Universe: NASA and the Development of Astrobiology* (Rutgers Univ. Press, 2004).
4. <http://blogs.discovermagazine.com/loom/category/arsenic-life/>
5. Wolfe-Simon, F. *et al. Science* **332**, 1163–1166 (2011).
6. Reaves, M. L., Sinha, S., Rabinowitz, J. D., Kruglyak, L. & Redfield, R. J. *Science* **337**, 470–473 (2012).
7. Erb, T. J., Kiefer, P., Hattendorf, B., Günther, D. & Vorholt, J. A. *Science* **337**, 467–470 (2012).
8. Simpson, G. G. *Science* **143**, 769–775 (1964).
9. Bada, J. L. *Science* **307**, 46 (2005).
10. Chyba, C. F. *Science* **308**, 495–496 (2005).
11. Knoll, A. H., Canfield, D. E. & Konhauser, K. O. in *Fundamentals of Geobiology* 1–4 (Wiley-Blackwell, 2012).
12. Kerr, R. A. *Science* **336**, 1500–1503 (2012).
13. Kivelson, M. G. *et al. Science* **289**, 1340–1343 (2000).

Facing up to complexity

The anatomy of three new fossils, including a face, lends support to the hypothesis that there were at least two parallel lineages early in the evolutionary history of our own genus, *Homo*. **SEE LETTER P.201**

BERNARD WOOD

The task of palaeoanthropologists is to reconstruct the evolutionary history of the period between our species, *Homo sapiens*, and the ancestral species we share exclusively with chimpanzees and bonobos. There must have been a ladder-like sequence of species connecting us with that common ancestor; but it is unclear whether our section of the 'tree of life' is restricted to this ancestor-descendant sequence, or whether it includes other, now extinct, lineages. Might there have been multiple lineages early in the history of our own genus, *Homo*? In this issue, Meave Leakey *et al.*¹ (page 201) describe fossils recovered from Koobi Fora in northern Kenya that provide compelling evidence for at least two *Homo* lineages as early as 2 million years ago.

For the first half of the last century, conventional wisdom was that the most primitive member of the genus *Homo* was *Homo erectus*, best known from fossils found at sites in China and southeast Asia. These creatures had small heads with prominent brows, and what little is known of their limbs suggests that they stood upright and walked on two feet in much the same way as we do. Just over 50 years ago, a series of discoveries at Olduvai Gorge in Tanzania² prompted palaeontologist Louis Leakey and his colleagues³ to add an even more primitive species at the base of the genus *Homo* (Fig. 1), which they called *Homo habilis* to reflect the assumption that members of the species had fashioned the simple stone tools found in the same sediment layers at Olduvai.

The brain of *H. habilis* was smaller than that of *H. erectus*, and the rest of its skeleton was more ape-like⁴. Researchers at Koobi Fora subsequently found further evidence of *H. habilis*, but they also found other early *Homo* cranial fossils that did not so obviously resemble those from the Olduvai site⁵.

In the late 1970s, I was given the task of trying to make sense of the collection of early *Homo* fossils from Koobi Fora, which included braincases, faces, lower jaws and teeth. Although a strong case had been made that the variation among the *H. habilis* fossils from Olduvai could be subsumed within a single species⁶, the question remained: did the two forms of early *Homo* found at Koobi Fora come from males and females of a single species, or did they represent two species? It was not easy to accommodate the variation among

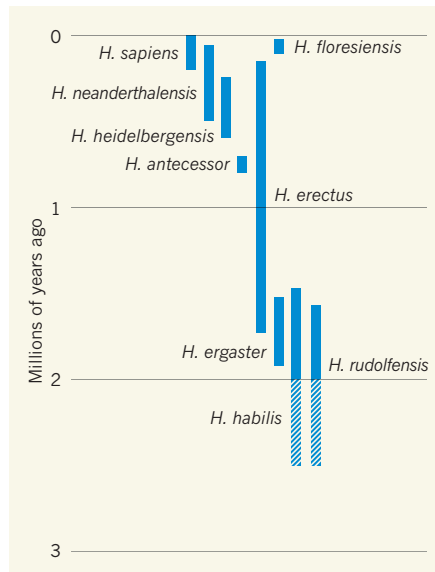


Figure 1 | Branching to humans. The genus *Homo*, to which modern humans (*Homo sapiens*) and several extinct species belong, most probably arose between 2 million and 3 million years ago, but fossil evidence before 2 million years ago (hatched lines) is frustratingly sparse. New fossil discoveries from Leakey and colleagues¹ strengthen the case for recognizing at least two evolutionary lineages at the early stages of the genus, to which the species *Homo habilis* and *Homo rudolfensis* are assigned. However, whether either of these two lineages was ancestral to *Homo erectus*, let alone to modern humans, remains uncertain.

the Koobi Fora early *Homo* fossils within one species. There were two distinct facial morphologies, and although the larger faces could have belonged to males, and the smaller ones to females, of the same species, the differences were unlike those we see between the males and females of modern humans and the living apes⁷. This discrepancy led researchers to recognize a second early *Homo* species, *Homo rudolfensis*⁸.

However, the existence of a second early *Homo* species remained controversial, as there was no single *H. rudolfensis* fossil specimen that contained both the face and the lower jaw. It was also unfortunate that, although the type specimen (the specimen that is linked with that species name) of *H. habilis* (called OH 7) had a lower jaw that contained teeth, it lacked a face, whereas the type specimen of *H. rudolfensis* (KNM-ER 1470) had a face, but neither tooth crowns nor a lower jaw. I made the assumption that the big-boned face of *H. rudolfensis*

would be matched by a robust jaw and large chewing teeth, and ascribed lower jaws with these attributes (such as KNM-ER 1802 from Koobi Fora) to *H. rudolfensis*⁸.

Leakey and colleagues' three new specimens¹ test these taxonomic hypotheses. The fossils are a well-preserved face (KNM-ER 62000), a well-preserved lower jaw (KNM-ER 60000) and a fragmentary lower jaw (KNM-ER 62003). None of the three specimens is as old as KNM-ER 1470, which is approximately 2 million years (Myr) old; the face and the fragmentary lower jaw are between 1.95 and 1.9 Myr old, and the better-preserved lower jaw is younger still, at around 1.83 Myr old. In a nutshell, the anatomy of the specimens supports the hypothesis of multiple early *Homo* species⁹, but refutes the hypothesis that lower jaws like KNM-ER 1802 went with the type of face — KNM-ER 1470 — that belongs to *H. rudolfensis*.

The KNM-ER 62000 face, which belongs to a late-juvenile individual, is smaller than that of KNM-ER 1470, yet it has the same shape — in both, the cheek bones join the palate quite far forward. Unlike KNM-ER 1470, the KNM-ER 62000 face does contain a few teeth, and the authors describe its chewing teeth as moderately sized and mesiodistally long (referring to the size and shape of the tooth crowns). Its palate is distinctively short, and the first of the two premolar teeth is at the corner of the almost right angle that is formed between the flat anterior tooth row and the straight, short, posterior tooth row (see Fig. 3a of the paper¹). By contrast, in *H. habilis* the two tooth rows are more curved, and it is the canine that is at their more rounded intersection. The new lower jaws are a much better match to the distinctively shaped KNM-ER 1470 and KNM-ER 62000 palates than are mandibles such as KNM-ER 1802.

So where do we go from here? More work needs to be done using the faces and lower jaws of modern humans and great apes to check how different the shapes of the palate and lower jaw can be among individuals in living species. We also need to find a way to formally estimate the likelihood that the OH 7 lower jaw came from the same species as either KNM-ER 60000 or KNM-ER 1802. If the latter can be accommodated within *H. habilis*, then all well and good, but if not (which I think is more likely), then could KNM-ER 1802 and its ilk⁸ represent a third species? Finally, some researchers have suggested^{9,10} that evidence from the face and jaws of *H. habilis* and *H. rudolfensis*, plus what little fossil evidence we have of these species' other body regions, stretches the definition of the genus *Homo* too far. Perhaps these two taxa belonged to a different lineage from that from which *H. sapiens* arose? My prediction is that by 2064, 100 years after Leakey and colleagues' description of *H. habilis*³, researchers will view our current hypotheses about this phase of human evolution as remarkably simplistic. ■

Bernard Wood is in the Center for the Advanced Study of Hominid Paleobiology, Department of Anthropology, George Washington University, Washington DC 20052, USA.
e-mail: bernardawood@gmail.com

1. Leakey, M. G. *et al.* *Nature* **488**, 201–204 (2012).
2. Leakey, M. D. *Olduvai Gorge: Excavations in Beds I and II, 1960–1963* (Cambridge Univ. Press, 1971).
3. Leakey, L. S. B., Tobias, P. V. & Napier, J. R.

- Nature* **202**, 7–9 (1964).
4. Ruff, C. *Am. J. Phys. Anthropol.* **138**, 90–100 (2009).
5. Leakey, M. G. & Leakey, R. E. *Koobi Fora Research Project Vol. 1* (Clarendon, 1978).
6. Tobias, P. V. *Olduvai Gorge: The Skulls, Endocasts and Teeth of Homo habilis* (Cambridge Univ. Press, 1991).
7. Wood, B. *Koobi Fora Research Project Vol. 4* (Clarendon, 1991).
8. Wood, B. *Nature* **355**, 783–790 (1992).
9. Wood, B. & Collard, M. *Science* **284**, 65–71 (1999).
10. Wood, B. & Baker, J. *Annu. Rev. Ecol. Evol. Syst.* **42**, 47–69 (2011).

OPTICS

Gain and loss mixed in the same cauldron

The mathematical concept of parity–time symmetry, which was first introduced in field theories of quantum mechanics, has been demonstrated experimentally in a large-scale optical system. [SEE ARTICLE P167](#)

LUCA RAZZARI & ROBERTO MORANDOTTI

The search for artificial structures exhibiting optical properties not found in nature is a major scientific endeavour. This effort has led, for example, to the development of photonic crystals¹ for controlling the propagation of electromagnetic waves by creating allowed and forbidden energy bands. It has also led to the design of metamaterials aimed at rendering objects invisible or at allowing super-resolution imaging^{2,3}. Yet, irrespective of the specific system involved, most of these advances have so far been based on a judicious crafting of the spatial distribution of the refractive index across passive optical components, which require no input power to function. Indeed, refractive-index engineering is considered today to be one of the cornerstones of modern photonics. On page 167 of this issue, Regensburger and co-workers⁴ herald the next frontier in this research field.

The authors' proposal is based on the fact that the refractive index is, in the most general case, a complex quantity. Whereas the real part of the index is responsible for the bending and slowing down of light, the imaginary part can lead to either amplification (gain) or absorption (loss) of light within a material. By modulating both the real and the imaginary part of the index, the authors were able to produce the first example of a new class of optical structure. The structure displays the recently developed concept of parity–time (PT) symmetry — a property of physical systems that are invariant under time inversion and mirror reflection. Key to achieving this symmetry was the authors' deliberate use of both gain and loss. Interestingly,

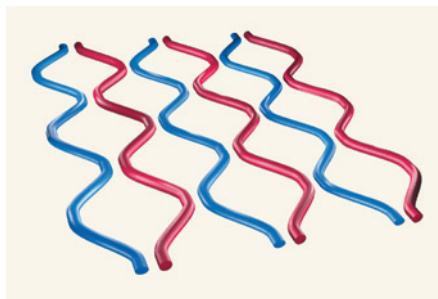


Figure 1 | A parity–time optical network. Regensburger *et al.*⁴ have designed an optical system that displays parity–time symmetry. The system is analogous to a spatially periodic network of channels that either amplify (red channels) or absorb (blue channels) light.

PT symmetry was originally introduced in field theories of quantum mechanics⁵.

Although the impact of PT symmetry in quantum mechanics is still open to debate, it has been recognized⁶ that this concept could flourish in the field of optics if gain, loss and refractive index are manipulated. For an extended optical system to be PT symmetric, the refractive-index profile needs to be symmetric in its real part and anti-symmetric in the imaginary one. So far, the implementation of such PT-symmetric devices has been hampered by technical difficulties, and has thus been limited to simplified proof-of-principle structures involving only two components^{7,8}. Regensburger and colleagues' work represents the first experimental demonstration of a large-scale PT-symmetric lattice, which introduces PT synthetic devices into the larger family of artificial optical systems.

To realize their artificial structure, and



50 Years Ago

The first practical research tool based on solid state lasers to be made available in the United Kingdom is announced by Kollsman Instrument, Ltd. The 'Pisto Laser', so called because of its double pistol-grip triggering, was designed and developed by the Company's associate in the United States of America, the Kollsman Instrument Corporation. It emits a beam of intense, coherent red light of wavelength 6929 Å., with a beam divergence of less than 0.3 degree ... The radiation output of the pulsed, ruby, laser beam is more than 10⁹ times the corresponding output of the Sun within the same frequency band ... In communications research the 'Pisto Laser' can be used to study the characteristics of very narrow beam-width transmission of optical data; in medical research, high-intensity cauterization; in crystallography, crystal structures using non-linear light transmission; in photography, pictures of extreme clarity; and in biology, biological responses to high-intensity monochromatic stimulation.

From *Nature* 11 August 1962

100 Years Ago

A good deal has been recently heard about "holes in the air" in connection with sudden collapses of flying machines. Prof. W. J. Humphreys, of the Washington Weather Bureau, writing in *The Popular Science Monthly* for July, classifies the eight different types of atmospheric disturbance as follows:— A vertical group, including aerial fountains, aerial cataracts, aerial cascades, and aerial breakers, and a horizontal group, including wind layers, wind billows, and aerial torrents; in addition wind eddies fall under both groups. Holes in the sense of vacuous regions do not exist.

From *Nature* 8 August 1912

to study the transport of light therein, Regensburger *et al.* used an elegant experimental arrangement that operates in the temporal domain. The authors injected a sequence of light pulses into two connected optical-fibre loops that were designed to exhibit PT symmetry: the required anti-symmetry of the imaginary part of the refractive-index profile was attained by alternating gain and loss in the two loops using optical amplifiers and amplitude modulators; and the symmetry of the real component of the profile was introduced using phase modulators (devices that control where a light wave's peaks and troughs lie). Figure 1 shows a network that is a spatial equivalent of the authors' structure; each node of this spatial network corresponds to a specific 'time slot' of the temporal lattice.

Using this set-up, Regensburger *et al.* have observed unusual optical behaviour such as unidirectional invisibility. They demonstrate that their system can become totally invisible when light traverses it from one side, whereas it can still be seen when it is illuminated from the other. What's more, the authors suggest that the device could find interesting applications in laser science, in particular in the dynamical control of light power in laser cavities. Indeed, their set-up shares several similarities with a laser cavity known as an active mode-locking cavity, which produces a train of ultrashort optical pulses on the basis of active phase modulation and balance between gain and loss.

The demonstration of a large-scale PT-symmetric optical network in the temporal domain is undoubtedly of great importance. Yet the realization of a spatial analogue, such as a waveguide array or crystal, remains a challenge. Transferring PT-related concepts to the spatial domain would allow the implementation of new optical devices that could in principle be scaled down and arranged on a chip. In particular, we can envisage the use of PT symmetry in the next generation of metamaterials and plasmonic devices, which work by manipulating surface plasmons — collective, wave-like motions of free electrons on a metal surface. PT symmetry can put to good use these systems' light losses, which have always been considered problematic. Similarly, we can imagine devices that combine PT symmetry and nonlinear optical properties. Such composite systems could permit the exploration of unprecedented optical functions, further improving our ability to artificially manipulate light. ■

Luca Razzari and Roberto Morandotti
are at the INRS Center for Energy, Materials and Telecommunications, Varennes, Quebec J3X 1S2, Canada. L.R. is also in the Nanostructures Department, Istituto Italiano di Tecnologia, Genova, Italy.
e-mails: luca.razzari@iit.it;
morandot@emt.inrs.ca

1. Yablonovitch, E. *Phys. Rev. Lett.* **58**, 2059–2062 (1987).

2. Shalaev, V. M. *Nature Photon.* **1**, 41–48 (2007).
3. Pendry, J. B., Schurig, D. & Smith, D. R. *Science* **312**, 1780–1782 (2006).
4. Regensburger, A. *et al. Nature* **488**, 167–171 (2012).
5. Bender, C. M. & Boettcher, S. *Phys. Rev. Lett.*

- 80**, 5243–5246 (1998).
6. Makris, K. G., El-Ganainy, R., Christodoulides, D. N. & Musslimani, Z. H. *Phys. Rev. Lett.* **100**, 103904 (2008).
7. Guo, A. *et al. Phys. Rev. Lett.* **103**, 093902 (2009).
8. Rüter, C. E. *et al. Nature Phys.* **6**, 192–195 (2010).

ATMOSPHERIC CHEMISTRY

The X factor

Measurements in a forest reveal a previously unknown atmospheric oxidant that acts as a source of sulphuric acid — one of the main precursors for the formation and growth of aerosol particles and clouds. [SEE LETTER P. 193](#)

WAYNE HEARD

Background signals are a nuisance for atmospheric scientists doing fieldwork, because they limit the sensitivity and precision of the instruments used to measure atmospheric composition. But it pays to inspect background signals carefully, particularly if the molecule under study is the hydroxyl radical (OH[•], referred to here as OH for simplicity), the most important oxidant in the atmosphere. Just such an inspection led Mauldin *et al.*¹ to discover a previously unknown atmospheric oxidant, as they report on page 193 of this issue. Their findings should help to refine models of atmospheric oxidation processes.

The story begins in a Finnish forest (Fig. 1), where the authors were indeed measuring OH. Their method involved adding sulphur dioxide (SO₂) to airstream samples so that it reacts with OH to form sulphuric acid (H₂SO₄), which is then detected by a mass spectrometer. When they deliberately removed OH from their samples using a chemical scavenger, however, they noticed that the background signal was actually larger than the OH signal. In other words, something in the forest other than OH was converting sulphur dioxide into sulphuric acid in their analyses, and so must also have been doing so in the atmosphere above the forest.

The atmospheric concentration of the unknown oxidant — which Mauldin *et al.* dubbed 'X' — was found to exceed that of OH, most noticeably in the evenings and at night. The concentration of X showed no clear daily cycle, however, suggesting that it forms from the reaction of surface emissions, such as naturally produced hydrocarbons, with ozone (O₃).

To test this hypothesis, the authors performed laboratory experiments in which they exposed sulphur dioxide to mixtures of ozone and various unsaturated hydrocarbons (alkenes). The reactions of alkenes with ozone are known to produce OH, but the levels of sulphuric acid observed in the experiments were well above those that would have been expected from the reaction of OH with sulphur dioxide alone. This was especially true

when the authors reacted ozone with limonene and α -pinene, two alkenes emitted by trees². To prove beyond reasonable doubt that plant emissions are linked to X, Mauldin *et al.* went back to the forest and placed cut tree branches close to the inlet of their oxidant-measuring instrument. Sure enough, they observed substantial levels of X.

The authors propose that X is probably a stabilized Criegee intermediate³; such molecules are free radicals that form from the reaction of ozone with alkenes, and are known to react with sulphur dioxide⁴. But the rates of the reactions of Criegee intermediates with sulphur dioxide were thought to be too slow to have any atmospheric relevance to the formation of sulphuric acid⁵. So is the authors' interpretation correct?

Support comes from a paper published earlier this year⁶, in which the simplest Criegee intermediate, CH₂OO, was detected directly for the first time, and was shown to be much more reactive towards sulphur dioxide than previously thought. When Mauldin *et al.*¹ estimated the rate constants — measures of reaction rates — for reactions of sulphur dioxide with the Criegee intermediates generated from α -pinene and limonene, they found that these reactions, too, were faster than previously assumed.

In their field experiments, the authors were able to measure atmospheric concentrations of sulphuric acid at the same time as they detected OH. They therefore compared the concentration of atmospheric sulphuric acid in the Finnish forest with the concentration that would have been produced by the oxidation of sulphur dioxide by OH alone, which they calculated from the measured concentrations of atmospheric OH and sulphur dioxide. They observed a difference at all times of the day, with the difference scaling with the concentration of X, clearly connecting X to the formation of the acid.

By determining the rate constant for the reaction of X with sulphur dioxide in their laboratory experiments, Mauldin *et al.* calculated the concentration of sulphuric acid in



Figure 1 | Something in the air. This tower at the SMEAR II field station, Finland, supports several instruments that are used to analyse the atmosphere above the forest canopy. Mauldin *et al.*¹ have detected a new oxidant in the atmosphere above the station.

the atmosphere that was not derived from OH. When they added this value to the calculated concentration of acid that was derived from OH, they found that the total agreed well with the observed atmospheric concentration of the acid in the forest.

The technique⁷ used by Mauldin and colleagues to measure OH is known as chemical ionization mass spectroscopy (CIMS), and it has been used in a range of environments. It is therefore surprising that the significance of background signals has not been recognized in previous studies. That said, the forested environment studied by the authors produces large quantities of alkene emissions, and so provides ideal conditions for the formation of X. Measurements of X are now needed in other environments, to determine its global impact on the production of atmospheric sulphuric acid.

Mauldin *et al.* propose that X converts sulphur dioxide to sulphur trioxide (SO₃), which then reacts with water vapour to form sulphuric acid (see Fig. 3 of the paper¹). But sulphur trioxide might not be the only product of sulphur dioxide's reaction with X, and the authors do not determine the — possibly multistep — reaction mechanism for this transformation. Indeed, X might not be a Criegee intermediate at all; perhaps a derivative of it, or another compound, reacts with sulphur dioxide⁸. Direct identification and field measurements of X are necessary to resolve this issue.

Furthermore, the authors evaluate only the role of X in oxidizing sulphur dioxide to sulphuric acid. Until the rate constants for reactions of X with a wide range of atmospheric species have been determined, its overall importance for atmospheric chemistry relative to OH will remain unknown — even though its concentration in Mauldin and colleagues' study exceeds that of OH. A more practical issue is that, if X's contribution to the

production of sulphuric acid is greater than that of OH, it will make the measurement of OH by CIMS more challenging, because of the need to subtract a background signal larger than the OH signal.

The atmospheric oxidation of sulphur dioxide is closely associated with the rate of aerosol-particle formation and growth, and with the production of cloud condensation nuclei⁹ —

microscopic particles around which cloud droplets coalesce. In calculations predicting regional and global temperature rises caused by human activities, the largest uncertainties are associated with aerosols and clouds¹⁰. Until now, OH has been assumed to be the only oxidizer that converts sulphur dioxide to sulphuric acid. Mauldin and colleagues' findings will therefore help to reduce the uncertainties in climate predictions that aim to take into account future changes in man-made sulphur dioxide emissions and in natural hydrocarbon emissions from plants. ■

Dwayne Heard is at the School of Chemistry, University of Leeds, Leeds LS2 9JT, UK.
e-mail: d.e.heard@leeds.ac.uk

1. Mauldin, R. L. *et al.* *Nature* **488**, 193–196 (2012).
2. Guenther, A. *et al.* *J. Geophys. Res.* **100**, 8873–8892 (1995).
3. Criegee, R. *Angew. Chem. Int. Edn Engl.* **14**, 745–752 (1975).
4. Hatakeyama, S. *et al.* *J. Phys. Chem.* **90**, 4131–4135 (1986).
5. Johnson, D. *et al.* *J. Phys. Chem. A* **105**, 2933–2935 (2001).
6. Welz, O. *et al.* *Science* **355**, 204–207 (2012).
7. Eisele, F. L. & Tanner, D. J. *J. Geophys. Res.* **96**, 9295 (1991).
8. Cox, R. A. & Penkett, S. A. *Nature* **230**, 321–322 (1971).
9. Sipilä, M. *et al.* *Science* **327**, 1243–1246 (2010).
10. Solomon, S. D. *et al.* (eds) *The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2007).

STRUCTURAL BIOLOGY

Dynamic binding

Nuclear magnetic resonance spectroscopy has been used to establish a vital role for protein motion in the formation of a protein–DNA complex. The finding potentially opens up fresh approaches for modifying protein function. SEE LETTER P.236

ANDREW J. BALDWIN & LEWIS E. KAY

The continuing development of the tools of structural biology and their increasingly sophisticated application to studies of a wide range of biological molecules are some of the most noteworthy accomplishments of biophysics. The structures obtained have been used to explain molecular function, to design and modify proteins so as to engineer new biological properties, and in the rational generation of pharmaceuticals. Yet despite the well-documented successes, there have also been many cases in which the beautiful, high-resolution pictures produced leave many questions unanswered. Part of the reason is that the biological function of a molecule is driven by free energy changes that depend on contributions from both static (enthalpic) and dynamic (entropic) interactions¹. Although

static structures can provide atomic-resolution information about enthalpy, describing entropy at a similar level of detail has been far more difficult.

On page 236 of this issue, Tzeng and Kalodimos² quantify the role of dynamics in their study of the catabolite activator protein, CAP. They have used nuclear magnetic resonance (NMR) spectroscopy to demonstrate that the binding activity of CAP can be regulated by conformational entropy — the entropy component associated with the number of conformational states that the protein samples.

CAP acts as a transcriptional activator — it binds to DNA to increase gene transcription. It is also an allosteric protein: binding of a ligand molecule at one site can introduce changes in both the structure^{3,4} and the dynamics⁵ of distal sites. More specifically, binding of a small nucleotide molecule, cyclic AMP (cAMP),

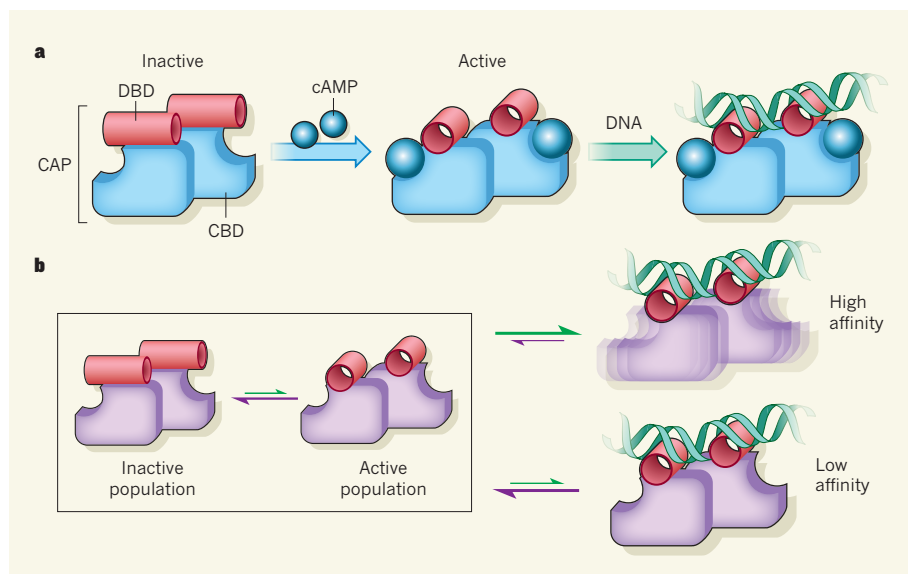


Figure 1 | Conformational entropy can modulate protein–DNA binding. **a**, When cAMP molecules bind to the inactive state of the dimeric catabolite activator protein (CAP) at specific sites in the cAMP-binding domains (CBDs), DNA-binding domains (DBDs) in the protein alter their orientation. This activates CAP so that it can bind with high affinity to DNA. **b**, Tzeng and Kalodimos² studied DNA binding to several engineered mutants of CAP, and found that conformational entropy can drive binding. In the CBD variant shown (purple), the equilibrium between inactive and active conformations is highly skewed towards the inactive conformation. Neglecting other factors, a higher affinity for DNA binding is realized if the dynamics of the CAP–DNA complex are increased (that is, if the complex has a large conformational entropy) relative to a complex that shows little change in motion upon binding.

to the cAMP-binding domain of CAP leads to substantial structural rearrangements in distal DNA-binding domains, priming the protein for DNA binding (Fig. 1a). Tzeng and Kalodimos studied CAP in the unbound and DNA-bound states.

The authors exploited CAP's allostery by engineering mutations in the protein at sites remote from the DNA-binding interface, but which nevertheless modulate DNA binding. By using NMR to study the derived mutants, as well as different nucleotide-bound forms of the protein, the authors established that the protein interconverts between inactive states that cannot bind DNA and active states that can, and that, for the mutants examined, the relative populations of these states can be very different.

In the simplest of binding models for CAP, the affinity of the protein for its target DNA is directly proportional to the fraction of molecules in the active conformation, as has been seen previously in different contexts for other systems^{6–9}. But Tzeng and Kalodimos observed little such correlation. Indeed, the affinities of some of the CAP mutants for DNA are 50-fold greater, and others are 25-fold lower, than would have been predicted on the basis of the numbers of molecules populating the active state. This clearly indicates that static structures alone cannot explain CAP's behaviour.

The authors therefore went on to measure the enthalpic and entropic contributions to the CAP mutants' DNA-binding affinities using a calorimetric technique. Although the resulting

data are informative, they are not at atomic resolution and they lump together contributions from a variety of terms. Of these contributions, one of the most useful to evaluate is conformational entropy, which counts the number of states adopted by bonds in CAP's 'backbone' and amino-acid side chains.

To gain more insight into the conformational entropy of the CAP mutants, Tzeng and Kalodimos carried out NMR experiments to quantify the amplitudes of motion of methyl groups in side chains, at the picosecond-to-nanosecond timescale (one picosecond is 10^{-12} seconds). They found that, on DNA binding, some of the mutants undergo very large net changes in conformational entropy that significantly increase the strength of association (Fig. 1b). In other cases, they observed that enthalpy changes drive binding, and that entropy changes oppose it.

Notably, the authors obtained an atomic-level description of how the picosecond-to-nanosecond dynamics of CAP respond to formation of the CAP–DNA complex from which the conformational entropy change for CAP–DNA binding was established. The pattern that emerges is not simple. For example, large changes in amplitudes of motion might have been expected only at regions close to the binding interface, but such changes extend much farther away, involving methyl groups more than 50 Å from the interface. Remarkably, the authors found a strong linear correlation between NMR-derived measures of conformational entropy change and the total

change in entropy measured by calorimetry. Such a correlation has previously been observed¹⁰ in the binding of the calmodulin protein to its target peptides.

Tzeng and Kalodimos's findings show that, although entropy changes are crucial for CAP's function, there are other ways of modulating the strength of DNA binding to the protein. Even though the correlation between the fractional population of CAP molecules in the active conformation and DNA-binding affinity is poor, increasing the population that is in the active state still remains one avenue for increasing affinity. The authors' NMR experiments revealed that, of the nine CAP mutants that occupy predominantly inactive structures, two interchange with active conformations on the millisecond timescale; approximately 2% of the molecules are in the active conformation for one of these mutants, and 7% for the other. Unlike the other seven variants, for which active populations could not be detected, these two mutants bound DNA.

Most interestingly, Tzeng and Kalodimos' work suggests an approach for manipulating protein function through allostery. For example, one could imagine targeting a drug to a site in an enzyme that is far removed from the active site, in order to modify the enzyme's function. By combining allosteric drugs of this sort with more traditional pharmaceuticals that bind to active sites, it may be possible to moderate the drug resistance that so often plagues conventional therapies. Understanding the fundamental roles of dynamics in protein function will also facilitate new ways of exploiting proteins and of modifying their activities. Tzeng and Kalodimos' work takes an important step in this direction. ■

Andrew J. Baldwin and Lewis E. Kay are in the Departments of Molecular Genetics, Biochemistry, and Chemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada. L.E.K. is also at the Hospital for Sick Children, Toronto.
e-mail: kay@pound.med.utoronto.ca

1. Karplus, M. & Kuriyan, J. *Proc. Natl Acad. Sci. USA* **102**, 6679–6685 (2005).
2. Tzeng, S.-R. & Kalodimos, C. G. *Nature* **488**, 236–240 (2012).
3. Popovych, N., Tzeng, S. R., Tonelli, M., Ebright, R. H. & Kalodimos, C. G. *Proc. Natl Acad. Sci. USA* **106**, 6927–6932 (2009).
4. Sharma, H., Yu, S., Kong, J., Wang, J. & Steitz, T. A. *Proc. Natl Acad. Sci. USA* **106**, 16604–16609 (2009).
5. Popovych, N., Sun, S., Ebright, R. H. & Kalodimos, C. G. *Nature Struct. Mol. Biol.* **13**, 831–838 (2006).
6. Li, P., Martins, I. R., Amarasinghe, G. K. & Rosen, M. K. *Nature Struct. Mol. Biol.* **15**, 613–618 (2008).
7. Volkman, B. F., Lipson, D., Wemmer, D. E. & Kern, D. *Science* **291**, 2429–2433 (2001).
8. Masterson, L. R. *et al. J. Mol. Biol.* **412**, 155–164 (2011).
9. Selvaratnam, R., Chowdhury, S., VanSchouwen, B. & Melacini, G. *Proc. Natl Acad. Sci. USA* **108**, 6133–6138 (2011).
10. Frederick, K. K., Marlow, M. S., Valentine, K. G. & Wand, A. J. *Nature* **448**, 325–329 (2007).

Parity–time synthetic photonic lattices

Alois Regensburger^{1,2}, Christoph Bersch^{1,2}, Mohammad–Ali Miri³, Georgy Onishchukov², Demetrios N. Christodoulides³ & Ulf Peschel¹

The development of new artificial structures and materials is today one of the major research challenges in optics. In most studies so far, the design of such structures has been based on the judicious manipulation of their refractive index properties. Recently, the prospect of simultaneously using gain and loss was suggested as a new way of achieving optical behaviour that is at present unattainable with standard arrangements. What facilitated these quests is the recently developed notion of ‘parity–time symmetry’ in optical systems, which allows a controlled interplay between gain and loss. Here we report the experimental observation of light transport in large–scale temporal lattices that are parity–time symmetric. In addition, we demonstrate that periodic structures respecting this symmetry can act as unidirectional invisible media when operated near their exceptional points. Our experimental results represent a step in the application of concepts from parity–time symmetry to a new generation of multifunctional optical devices and networks.

In designing an optical system, nature demands that only a few basic ‘ingredients’ be used: refractive index, gain and loss. There is no doubt as to how useful index contrast is in controlling optical dynamics. Shaping the refractive index profile has led to unprecedented advancements in optics ranging from the design and fabrication of photonic crystals¹ and photonic crystal fibres² to the exploration of nanoplasmonics^{3–6} and negative-index metamaterials⁷. Loss is abundant in physical systems, but is typically considered a problem. Gain, however, as afforded by lasers, is valuable in optoelectronics because it provides a means to induce lasing or to overcome losses^{4–6}. The question naturally arises as to whether new artificial optical structures and materials can be synthesized by mixing together these three ingredients in roughly comparable proportions, and, if so, how this can be done without running into uncontrollably growing or decaying optical modes. The answer may come from some recent abstract developments in quantum field theories.

In 1998, it was shown that a wide class of non-Hermitian Hamiltonians can actually possess entirely real spectra as long as they respect parity–time symmetry⁸. This is clearly counterintuitive given the fact that this symmetry is commonly associated with purely Hermitian systems. In quantum mechanics, the action of the parity operator \hat{P} is defined by the relations $\hat{x} \rightarrow -\hat{x}$ and $\hat{p} \rightarrow -\hat{p}$, and that of the \hat{T} operator leads to $\hat{p} \rightarrow -\hat{p}$, $i \rightarrow -i$ and $\hat{x} \rightarrow \hat{x}$, where \hat{x} and \hat{p} represent the position and momentum operators, respectively. In general, a Hamiltonian $\hat{H} = \hat{p}^2/2 + V(\hat{x})$ associated with a complex potential $V(\hat{x})$ is parity–time symmetric provided that it commutes with the parity–time operator. In this case, $\hat{H}\hat{P}\hat{T} = \hat{P}\hat{T}\hat{H}$ and, thus, \hat{H} and $\hat{P}\hat{T}$ may share a common set of eigenfunctions^{8–14}.

Given that the evolution of the system is described by a Schrödinger evolution equation of the form $i\partial_t\Psi = \hat{H}\Psi$, where $\partial_t\Psi$ denotes the partial derivative of Ψ with respect to time, it can be shown⁹ that a necessary (but not sufficient) condition for this to occur is that the complex potential involved in such a Hamiltonian satisfies $V(x) = V^*(-x)$. This implies that the real part of the potential must be an even function of position and that the imaginary part must be odd. In such pseudo-Hermitian configurations, the eigenfunctions are no longer orthogonal and, hence, the vector space of eigenmodes is skewed^{8,10}. Even more intriguing is the possibility of a sharp, symmetry-breaking transition once a non-hermiticity parameter exceeds

a certain critical value, the ‘parity–time threshold’. In such a regime, the Hamiltonian and the parity–time operator no longer have the same set of eigenfunctions (even though they commute) and as a result the eigenvalues of the system cease to be all real. In addition, this broken parity–time symmetry phase is associated with the appearance of exceptional points^{8–11}, where the eigenvalue branches merge and parity–time symmetry breaks down.

Although the ramifications of these developments are still a matter of debate within the framework of theoretical physics, it has recently been recognized that optics can provide a productive test bed where the notions of parity–time symmetry can be experimentally explored, and ultimately used^{15–17}. Given that such photonic systems are entirely classical, they can be fully realized without introducing any conflict with the hermiticity of quantum mechanics. In optics, a complex potential can be readily built by judiciously incorporating, in a balanced way, regions having optical gain and loss¹⁵. What allows this duality between quantum mechanics and optics is the isomorphic nature of the wave equations involved (Supplementary Methods, section 1.9). In this case, the complex refractive index plays the part of a potential and ‘parity–time symmetry’, hereafter PT symmetry in the context of optics, therefore demands^{15–17} that $n(x) = n^*(-x)$. However, so far, the realization of structures that simultaneously exhibit a symmetric refractive index distribution and an antisymmetric gain/loss profile has been hampered by technical difficulties. As a result, PT symmetry has been observed only in elemental two-component systems^{18–21}. It will be important to devise new, versatile platforms where such phenomena can be explored. If successful, a new generation of optical devices, materials and networks may result, including, for example, unidirectional on-chip devices²² as well as PT-symmetric, high-power laser systems and laser oscillators^{23–31}. Finally, ideas from PT symmetry may provide a viable route to overcoming losses that have so far hindered progress in other areas of applied physics such as plasmonics and metamaterials^{5–7}.

In this Article, we report the experimental demonstration of temporally resolved optical beam dynamics in large-scale, PT-symmetric lattices. The unusual band structure associated with such extended systems is probed along with power unfolding and secondary emissions resulting from the skewed character of the Floquet–Bloch modes. In addition, we observe the unidirectional invisibility of a

¹Institute of Optics, Information and Photonics, University of Erlangen–Nürnberg, Staudtstraße 7/B2, 91058 Erlangen, Germany. ²Max Planck Institute for the Science of Light, Günther-Scharowsky-Straße 1, Bau 24, 91058 Erlangen, Germany. ³CREOL, College of Optics and Photonics, University of Central Florida, Orlando, Florida 32816–2700, USA.

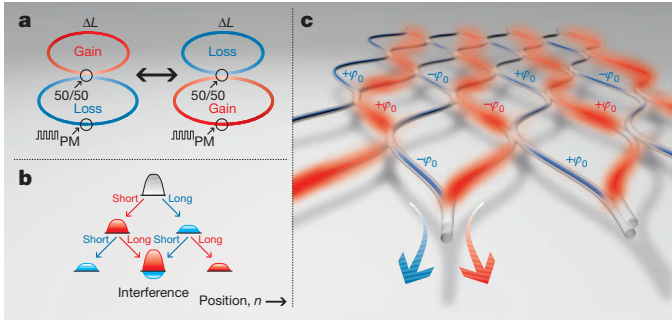


Figure 1 | PT-symmetric fibre networks. **a**, Two coupled fibre loops periodically switching between gain and loss as used in the experiment (Supplementary Fig. 1). Pulses are delayed or advanced as a result of a length difference ΔL between the loops. PM, phase modulator. **b**, Pulse evolution in the networks. Passages through short and long loops are indicated. **c**, Equivalent PT-symmetric lattice network. Gain (red) and loss (blue) channels are positioned antisymmetrically and are periodically coupled. Moreover, the real part of the potential is evenly imposed by phase shifts $\pm\varphi_0$.

PT-symmetric temporal grating when operating close to the system's exceptional points. Such structures have no effect on transmitted light, and their reflectivity greatly varies with the direction of propagation. The underlying flexibility of the arrangements proposed can be used to implement a diverse family of 'PT synthetic' structures and networks with altogether new properties and functionalities.

In our experiments, large-scale, PT-symmetric optical networks are realized in the temporal domain. This is accomplished by shaping the evolution of a sequence of pulses in two appropriately designed fibre loops connected by a 50/50 coupler (Fig. 1a and Supplementary Methods, section 1). A length differential ΔL between the two pathways leads to discretized arrival times for the optical pulses and to a transverse coupling between neighbouring time slots. The induced transverse pulse transport^{32,33} can be directly observed by monitoring

the output of tap nodes. A pulse traversing the short loop will be advanced, whereas pulses propagating along the long loop will be temporally delayed. The resulting pulse trains emerging from the short and long loops eventually interfere at the coupler in a manner depending on the phases acquired along their respective pathways (Fig. 1b and Supplementary Methods).

In this system, PT symmetry is imposed by temporally alternating gain and loss in the two loops by means of optical amplifiers in conjunction with amplitude modulators. In addition, the even, real part of the optical PT potential ($\text{Re}\{n(x)\}$) can be discretely introduced using phase modulation. The PT nature of this configuration becomes apparent by considering the fully equivalent spatial waveguide network shown in Fig. 1c, where the time slots associated with the set-up described here are mapped on discrete transverse sites n . A pulse delay or advancement in the temporal domain corresponds to a transition towards the left or the right in the spatial network of Fig. 1c. The light evolution in these spatiotemporal lattices is described by the following recursion equations³³:

$$\begin{aligned} u_n^{m+1} &= \frac{G^{\pm 1/2}}{\sqrt{2}} (u_{n+1}^m + i v_{n+1}^m) \\ v_n^{m+1} &= \frac{G^{\mp 1/2}}{\sqrt{2}} (i u_{n-1}^m + v_{n-1}^m) e^{i\varphi(n)} \end{aligned} \quad (1)$$

In the temporal domain, u_n^m and v_n^m denote the amplitudes of a sequence of pulses occupying time slots n within the short and long loop, respectively, after the m th round trip. The phase function $\varphi(n)$ provides the symmetric, real part of the PT potential, by imposing a phase $\pm\varphi_0$ on the pulses, and the antisymmetric, imaginary component of the effective potential is induced by a gain/loss factor G . In equation (1), the exponent of G switches from $-1/2$ to $+1/2$ between alternate loops in every step m ; that is, the loops are repeatedly switched between gain and loss in equal amounts.

Our experimental set-up (Fig. 1a), as well as its equivalent optical network (Fig. 1c), are periodic in both m and n , thus leading to a band

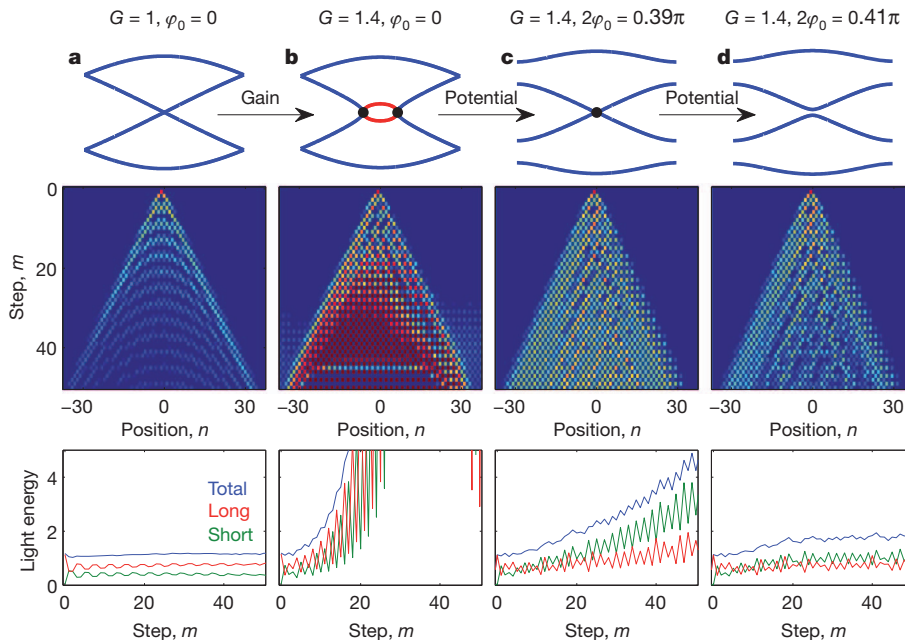


Figure 2 | Band structure and measured single-pulse evolution in the PT synthetic network. **a**, Passive system ($G = 1$) without phase potential ($\varphi_0 = 0$) with an entirely real (blue) band structure and no band gaps. The evolution of a pulse injected into the long loop is similar³³ to a quantum walk³². Measured pulse intensities at position n and step (loop round trip) m are indicated in the logarithmic colour scale: red, high intensity (1); dark blue, low intensity ($10^{-2.05}$). **b**, Same as **a**, but with net gain/loss ($G = 1.4$). Parts of the band structure become imaginary (red), leading to an exponential energy growth.

Black points in the figures denote exceptional points. **c**, Same as **b**, but with a phase potential ($2\varphi_0 = 0.39\pi$) partly stabilizing the system, leading to a linear growth in energy. **d**, A further increase of the real-valued potential ($2\varphi_0 = 0.41\pi$) opens a band gap, making the band structure again entirely real. Power unfolding is observed in this case; that is, the optical energy is no longer conserved but oscillates about a mean value because the eigenmodes are not orthogonal. For data in both loops, colour scales and a comparison with simulations, see Supplementary Figs 4 and 5.

structure as shown in Fig. 2 (see Supplementary Fig. 6 for details). In fact, all the unusual dynamics in these PT synthetic networks can be traced back to their peculiar band features. As Fig. 2 indicates, the bands can show new characteristics including the formation of exceptional points and the possibility of band-merging effects^{15,17}—both a direct outcome of a spontaneously broken PT symmetry. In the passive case, when the imaginary part of the potential is absent³³, two bands form without a gap (Fig. 2a), and the system hence has no PT threshold. Therefore, even a small amount of gain/loss G abruptly forces the bands to merge at the exceptional points, where a transition to imaginary eigenvalues occurs (Fig. 2b). To establish a finite threshold, a symmetric phase potential $\varphi(n)$ has to be introduced into the time lattice. The presence of such a potential forces the bands apart, thus creating a band gap. In this arrangement, the spectrum is again entirely real, in spite of the fact that the system is not Hermitian (Fig. 2d).

In the experiment, we probe the entire band structure of this lattice by simultaneously exciting all bands with a single pulse. Although power is conserved in the passive time lattice (Fig. 2a), exponential growth occurs (Fig. 2b) above the point at which PT symmetry is broken, where the spectrum ceases to be real. However, if the system is operated at threshold, with only exceptional points present (Fig. 2b, c, black points), the power increases linearly in time as clearly shown in Fig. 2c, in agreement with previous theoretical predictions^{34–36}. Even below the symmetry-breaking point, power oscillations are observed (Fig. 2d). The question is why this occurs, given that the bands are entirely real. The answer has to do with the very fact that the network's Floquet–Bloch modes are no longer orthogonal and, consequently, power ‘hidden’ in the system can reappear during evolution^{15,35}. As a result, the total power in the system undergoes oscillations because of mode interference effects.

The versatility of the time-multiplexed lattice described here allows us to explore further the onset of PT symmetry breaking by performing a parameter scan throughout the accessible range of the phase potential φ_0 and gain/loss coefficient G (Fig. 3a). As Fig. 3a shows, if the phase modulation is strong enough we observe a harmonic co-existence between gain and loss (blue region) that can be suddenly broken by small parameter variations (as in the red area), resulting in an exponential energy rise as shown in Fig. 3b. However, in the exact

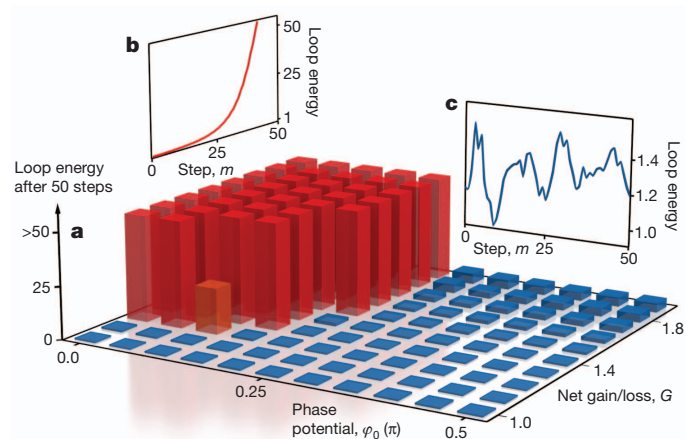


Figure 3 | Exploring the PT threshold of a lattice using a two-parameter scan. **a**, Total energy experimentally observed after 50 round trips m . A wide range of accessible parameters corresponding to the linear potential φ_0 and the magnitude of gain/loss G is experimentally searched, revealing the PT phase transition boundary, within which a sudden jump in energy occurs (red region). The small energy increase observed outside this region is caused by experimental imperfections and power unfolding. **b**, **c**, Observed evolution in the total energy of the system for partly complex (**b**; see Fig. 2b) and entirely real (**c**; see Fig. 2d) lattice spectra: unlimited exponential growth versus power oscillations.

PT phase the power oscillates during evolution (Fig. 3c). As both gain/loss G and the phase potential $\varphi(n)$ can be dynamically changed during the course of light propagation, it is possible to implement abrupt and gradual transitions from PT-symmetric to passive regimes in our temporal photonic lattice. Moreover, the sharp transitions between these phases can provide a precise mechanism for dynamic power control in laser cavities. Given the conceptual similarity of our set-up (Fig. 1a) to figure-eight fibre lasers, it might be possible to apply our modulation schemes (Supplementary Fig. 3) to achieve enhanced pulse control. For example, for the phase in which PT symmetry is broken, pulse trains with a fixed phase relation and a spacing defined by the path difference and not the total length of the loops could be produced.

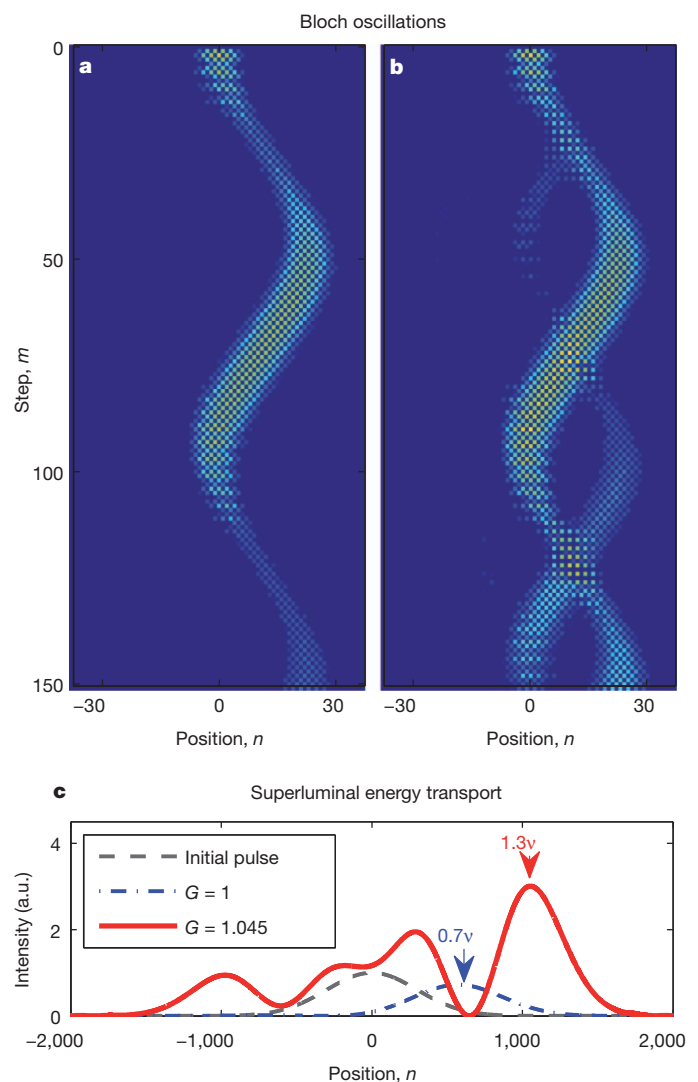


Figure 4 | Bloch oscillations (experiment) and superluminal energy transport (simulation). **a**, **b**, Bloch oscillations in the passive system (**a**) and in the system with gain/loss ($G = 1.08$) (**b**). A broad Gaussian input distribution is used to excite a narrow mode spectrum near the centre of the Brillouin zone. For each encounter with the imaginary parts of the band structure, a secondary light beam is emitted. A weak phase gradient $e^{i\alpha m}$ with $\alpha = \pi/25$ is applied to the long loop. Logarithmic colour scale: red, high intensity (1); dark blue, low intensity ($10^{-1.8}$). **c**, A broad initial pulse excitation with a narrow frequency spectrum close to the exceptional point speeds up the energy transport into the superluminal regime. The field distribution after $m = 800$ steps is shown for the passive case ($G = 1$) and the PT case ($G = 1.045$). Here v is the maximum possible speed of excitation spreading in the passive case (one step per round trip m). a.u., arbitrary units.

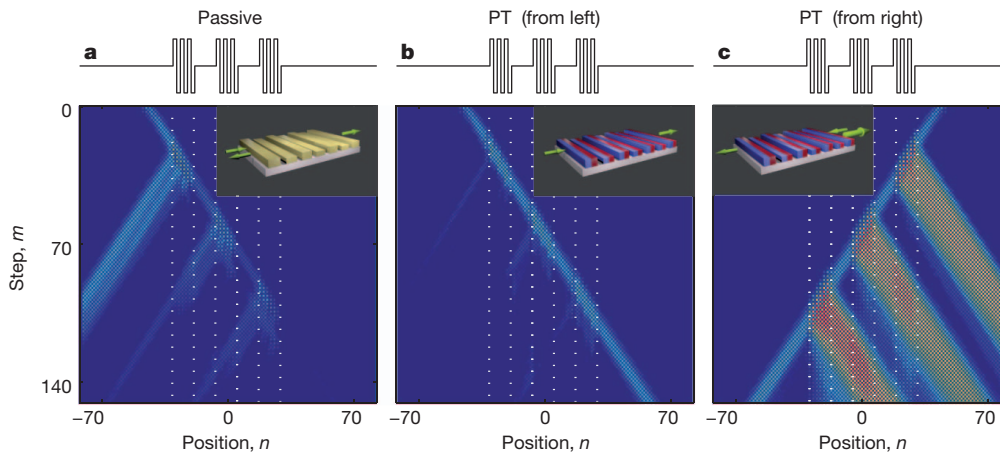


Figure 5 | Experimentally observed unidirectional invisibility of PT-symmetric Bragg scatterers. **a**, A broad pulse sequence is scattered off a series of passive Bragg elements ($2\phi_0 = 0.36\pi$), which are implemented via phase modulation. **b**, Adding antisymmetric gain/loss ($G = 1.3$) at the positions of the scatterers puts the system at the PT threshold (PT). This renders the gratings invisible to a beam impinging from the left. **c**, If light comes from the right or,

Pronounced outbursts of radiation are also observed when a broad input is subjected to a weak phase gradient in a PT-symmetric network above the symmetry-breaking point. As the beam dynamically changes its tilt, it performs Bloch oscillations^{37–39} and sweeps the whole band structure (Fig. 4a, b). In this case, every time the field experiences imaginary eigenvalues, a secondary beam is emitted and the total power is amplified (Fig. 4b). Finally, the merging of the two bands at the exceptional points has even more intriguing consequences. Numerical simulations show that as the slope of the bands tends to infinity, the associated group velocity of a narrow-bandwidth wave packet increases and can even exceed the speed of light in the fibre⁴⁰ (without violating causality^{41,42}), as indicated in Fig. 4c. This anomalous speed of the peak intensity is made possible by a gain-assisted growth of the distribution's leading tail. We note that all these features are a direct outcome of pseudo-hermiticity and have no analogue whatsoever in passive configurations.

Finally, we performed scattering experiments on a periodic, PT-symmetric temporal structure—a configuration analogous to a spatial grating (Fig. 5). As recently predicted^{43,44}, such structures have surprising behaviour such as unidirectional invisibility and unconventional reflection characteristics. More specifically, light propagating in such a system can experience reduced or enhanced reflections (with a coefficient that can even exceed one) depending on the direction of propagation. This is because left–right symmetry is broken in this PT network and propagation is no longer invariant when gain and loss are exchanged in time. Even more remarkable is what happens at the PT threshold: light waves entering the structure from one side do not experience any reflection and can fully traverse the grating with complete transmission. Given that this occurs without acquiring any phase imprint from the system, the periodic structure is essentially invisible⁴⁴. However, if light is incident from the opposite side, the coefficient of reflection exceeds one. In our set-up, we created a temporal Bragg scatterer by imposing a periodic phase modulation $\varphi(n)$ only within a finite time window. Optical pulses travelling outside this window do not experience this periodic potential. In the absence of any gain or loss, the resulting passive configuration acts on incoming light as a reflector, very much like a spatial Bragg stack (Fig. 5a). Gain and loss was subsequently added to the phase modulation in a PT-symmetric fashion. Figure 5 shows experimental results confirming these predictions. Both suppression of reflection and full transmission exactly at the symmetry-breaking point are clearly observed, suggesting invisibility of the scatterers.

equivalently, gain and loss are exchanged, reflection is strongly enhanced while transmission remains unchanged. We note that the scatterers are again invisible for the newly generated reflections returning from the left. The positions of the three scatterers are indicated by vertical dotted lines. Logarithmic colour scale: red, high intensity (1); dark blue, low intensity (10^{-2}).

Our results demonstrate that scalable PT synthetic discrete systems can be realized using building blocks that respect this reflection symmetry. The modular approach presented here can be easily extended to on-chip configurations⁴⁵, thus paving the way for the realization of PT synthetic devices and effective media with new and unexpected optical properties. Finally, similar concepts can be effectively used in other areas such as plasmonics and metamaterials, where a harmonic coexistence of gain and loss is ultimately required. These and related issues are now accessible experimentally using the platform described here.

METHODS SUMMARY

Operation of the time-multiplexed fibre network in its passive version is described in ref. 33 and its supplementary information. To introduce gain and loss into the network, an acousto-optic modulator (AOM) was inserted into each loop. Both AOMs were operated in the zeroth order to implement a variable attenuation from 0 to 6 dB without imposing a frequency shift on the signal. The switching time of the AOMs was fast enough to change the attenuation between subsequent round trips (or between subsequent positions n in the case of PT scattering). The net gain of both loops provided by semiconductor optical amplifiers was varied so that the optical energy remained constant when the AOM losses were set to 3 dB. This allows the effective gain/loss to be modulated by up to ± 3 dB (that is, $G = 2$) in each loop.

The phase modulation for the potential was provided by an electro-optic phase modulator in the long loop. The imposed phase function used was

$$\varphi(n) = \begin{cases} -\varphi_0 & \text{for } \text{mod}(n+3; 4) = 0 \text{ or } 1 \\ +\varphi_0 & \text{for } \text{mod}(n+3; 4) = 2 \text{ or } 3 \end{cases}$$

To observe PT Bloch oscillations, the phase factor $e^{i\varphi(n)}$ in equation (1) was replaced by a linearly increasing phase gradient e^{izm} .

Only every second position n can be accessed by the optical pulses in our loop network. Therefore, the amplitudes in between are set to zero. The colour scales in Figs 2, 4 and 5 are logarithmic and are shown in Supplementary Figures 4, 5, 8 and 9. Further details concerning the experimental procedures followed can be found in Supplementary Methods.

Received 1 March; accepted 8 June 2012.

- Joannopoulos, J. D., Villeneuve, P. R. & Fan, S. Photonic crystals: putting a new twist on light. *Nature* **386**, 143–149 (1997).
- Knight, J. C., Broeng, J., Birks, T. A. & Russell, P. St J. Photonic band gap guidance in optical fibers. *Science* **282**, 1476–1478 (1998).
- Barnes, W. L., Dereux, A. & Ebbesen, T. W. Surface plasmon subwavelength optics. *Nature* **424**, 824–830 (2003).
- Stockman, M. Spaser action, loss compensation, and stability in plasmonic systems with gain. *Phys. Rev. Lett.* **106**, 156802 (2011).

5. Berini, P. & De Leon, I. Surface plasmon–polariton amplifiers and lasers. *Nature Photon.* **6**, 16–24 (2011).
6. Noginov, M. A. *et al.* Demonstration of a spaser-based nanolaser. *Nature* **460**, 1110–1112 (2009).
7. Pendry, J. B. Negative refraction makes a perfect lens. *Phys. Rev. Lett.* **85**, 3966–3969 (2000).
8. Bender, C. M. & Boettcher, S. Real spectra in non-Hermitian Hamiltonians having PT symmetry. *Phys. Rev. Lett.* **80**, 5243–5246 (1998).
9. Bender, C. M. Making sense of non-Hermitian Hamiltonians. *Rep. Prog. Phys.* **70**, 947–1018 (2007).
10. Lévai, G. & Znojil, M. Systematic search for PT-symmetric potentials with real energy spectra. *J. Phys. Math. Gen.* **33**, 7165–7180 (2000).
11. Bender, C. M., Brody, D. C. & Jones, H. F. Complex extension of quantum mechanics. *Phys. Rev. Lett.* **89**, 270401 (2002).
12. Mostafazadeh, A. Pseudo-Hermiticity versus PT symmetry: the necessary condition for the reality of the spectrum of a non-Hermitian Hamiltonian. *J. Math. Phys.* **43**, 205–214 (2002).
13. Berry, M. V. Optical lattices with PT symmetry are not transparent. *J. Phys. A* **41**, 244007 (2008).
14. Ahmed, Z. Real and complex discrete eigenvalues in an exactly solvable one-dimensional complex PT invariant potential. *Phys. Lett. A* **282**, 343–348 (2001).
15. Makris, K. G., El-Ganainy, R. & Christodoulides, D. N. Beam dynamics in PT symmetric optical lattices. *Phys. Rev. Lett.* **100**, 103904 (2008).
16. El-Ganainy, R., Makris, K. G., Christodoulides, D. N. & Musslimani, Z. H. Theory of coupled optical PT-symmetric structures. *Opt. Lett.* **32**, 2632–2634 (2007).
17. Makris, K. G., El-Ganainy, R. & Christodoulides, D. N. PT-symmetric optical lattices. *Phys. Rev. A* **81**, 063807 (2010).
18. Rüter, C. E. *et al.* Observation of parity–time symmetry in optics. *Nature Phys.* **6**, 192–195 (2010).
19. Guo, A. *et al.* Observation of PT-symmetry breaking in complex optical potentials. *Phys. Rev. Lett.* **103**, 093902 (2009).
20. Schindler, J., Li, A., Zheng, M. C., Ellis, F. M. & Kottos, T. Experimental study of active LRC circuits with PT symmetries. *Phys. Rev. A* **84**, 040101 (2011).
21. Bittner, S. *et al.* PT symmetry and spontaneous symmetry breaking in a microwave billiard. *Phys. Rev. Lett.* **108**, 024101 (2012).
22. Ramezani, H., Kottos, T., El-Ganainy, R. & Christodoulides, D. Unidirectional nonlinear PT-symmetric optical structures. *Phys. Rev. A* **82**, 043803 (2010) CrossRef.
23. Chong, Y., Ge, L. & Stone, A. PT-symmetry breaking and laser-absorber modes in optical scattering systems. *Phys. Rev. Lett.* **106**, 093902 (2011).
24. Klaiman, S., Guenther, U. & Moiseyev, N. Visualization of branch points in PT-symmetric waveguides. *Phys. Rev. Lett.* **101**, 080402 (2008).
25. Miroshnichenko, A. E., Malomed, B. A. & Kivshar, Y. S. Nonlinearly PT-symmetric systems: spontaneous symmetry breaking and transmission resonances. *Phys. Rev. A* **84**, 012123 (2011).
26. Suchkov, S. V., Dmitriyev, S. V., Malomed, B. A. & Kivshar, Y. S. Wave scattering on a domain wall in a chain of PT-symmetric couplers. *Phys. Rev. A* **85**, 033825 (2012).
27. Uzdin, R. & Moiseyev, N. Scattering from a waveguide by cycling a non-Hermitian degeneracy. *Phys. Rev. A* **85**, 031804(R) (2012).
28. Kottos, T. Optical physics: broken symmetry makes light work. *Nature Phys.* **6**, 166–167 (2010).
29. Longhi, S. PT-symmetric laser absorber. *Phys. Rev. A* **82**, 031801(R) (2010).
30. Liertzer, M., Ge, L., Cerjan, A., Stone, A. D., Türeci, H. E., & Rotter, S. Pump-induced exceptional points in lasers. *Phys. Rev. Lett.* **108**, 173901 (2012).
31. Miri, M.-A. LiKamWa, P. & Christodoulides, D. N. Large area single-mode parity–time-symmetric laser amplifiers. *Opt. Lett.* **37**, 764–766 (2012).
32. Schreiber, A. *et al.* Photons walking the line: a quantum walk with adjustable coin operations. *Phys. Rev. Lett.* **104**, 050502 (2010).
33. Regensburger, A. *et al.* Photon propagation in a discrete fiber network: an interplay of coherence and losses. *Phys. Rev. Lett.* **107**, 233902 (2011).
34. Zheng, M. C., Christodoulides, D. N., Fleischmann, R. & Kottos, T. PT optical lattices and universality in beam dynamics. *Phys. Rev. A* **82**, 010103(R) (2010).
35. Graefe, E. M. & Jones, H. F. PT-symmetric sinusoidal optical lattices at the symmetry-breaking threshold. *Phys. Rev. A* **84**, 013818 (2011).
36. Scott, D. D. & Joglekar, Y. N. Degrees and signatures of broken PT symmetry in nonuniform lattices. *Phys. Rev. A* **83**, 050102 (2011).
37. Longhi, S. Bloch oscillations in complex crystals with PT symmetry. *Phys. Rev. Lett.* **103**, 123601 (2009).
38. Pertsch, T., Dannberg, P., Elflein, W., Bräuer, A. & Lederer, F. Optical Bloch oscillations in temperature tuned waveguide arrays. *Phys. Rev. Lett.* **83**, 4752–4755 (1999).
39. Morandotti, R., Peschel, U., Aitchison, J., Eisenberg, H. & Silberberg, Y. Experimental observation of linear and nonlinear optical Bloch oscillations. *Phys. Rev. Lett.* **83**, 4756–4759 (1999).
40. Szameit, A., Rechtsman, M. C., Bahat-Treidel, O. & Segev, M. PT-symmetry in honeycomb photonic lattices. *Phys. Rev. A* **84**, 021806 (2011).
41. Chiao, R. Superluminal (but causal) propagation of wave-packets in transparent media with inverted atomic populations. *Phys. Rev. A* **48**, R34–R37 (1993).
42. Wang, L. J., Kuzmich, A. & Dogariu, A. Gain-assisted superluminal light propagation. *Nature* **406**, 277–279 (2000).
43. Kulishov, M., Laniel, J. M., Bélanger, N., Azafia, J. & Plant, D. V. Nonreciprocal waveguide Bragg gratings. *Opt. Express* **13**, 3068–3078 (2005).
44. Lin, Z. *et al.* Unidirectional invisibility induced by PT-symmetric periodic structures. *Phys. Rev. Lett.* **106**, 213901 (2011).
45. Sansoni, L. *et al.* Two-particle bosonic-fermionic quantum walk via integrated photonics. *Phys. Rev. Lett.* **108**, 010502 (2012).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge financial support from DFG Forschergruppe 760, the Cluster of Excellence Engineering of Advanced Materials, SAOT and the German-Israeli Foundation. This work was also supported by NSF grant ECCS-1128520 and by AFOSR grant FA95501210148. Moreover, we thank J. Näger for technical support.

Author Contributions All authors contributed extensively to the work presented in this paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to U.P. (ulf.peschel@physik.uni-erlangen.de) and D.N.C. (demetri@creol.ucf.edu).

Deconstruction of a neural circuit for hunger

Deniz Atasoy¹, J. Nicholas Betley¹, Helen H. Su¹ & Scott M. Sternson¹

Hunger is a complex behavioural state that elicits intense food seeking and consumption. These behaviours are rapidly recapitulated by activation of starvation-sensitive AGRP neurons, which present an entry point for reverse-engineering neural circuits for hunger. Here we mapped synaptic interactions of AGRP neurons with multiple cell populations in mice and probed the contribution of these distinct circuits to feeding behaviour using optogenetic and pharmacogenetic techniques. An inhibitory circuit with paraventricular hypothalamus (PVH) neurons substantially accounted for acute AGRP neuron-evoked eating, whereas two other prominent circuits were insufficient. Within the PVH, we found that AGRP neurons target and inhibit oxytocin neurons, a small population that is selectively lost in Prader–Willi syndrome, a condition involving insatiable hunger. By developing strategies for evaluating molecularly defined circuits, we show that AGRP neuron suppression of oxytocin neurons is critical for evoked feeding. These experiments reveal a new neural circuit that regulates hunger state and pathways associated with overeating disorders.

Hunger involves interoceptive sensory neurons that monitor metabolic signals and consequently regulate food seeking and consumption behaviours. To isolate discrete neural circuit pathways controlling feeding behaviour, we have focused on neurons in the hypothalamic arcuate nucleus (ARC) that express the gene *agouti* related protein (*Agrp*). These are sensory neurons that are activated by circulating signals of energy deficit, such as ghrelin^{1–3}. Elevated AGRP neuron electrical activity rapidly evokes voracious eating, even in well-fed mice^{4,5}. Conversely, ablation of AGRP neurons results in aphagia⁶, and suppressing their electrical activity lowers food intake⁵. Therefore, these molecularly defined neurons that sense energetic need provide an entry point to neural circuits that mediate the ‘wisdom of the body’⁷ and are sufficient to orchestrate complex counter-regulatory behavioural responses.

AGRP axon projections reveal an anatomical map⁸ of brain regions that are potential downstream neuronal mediators of feeding behaviour (Supplementary Fig. 1). To establish pathways through which interoceptive sensory neurons acutely orchestrate feeding, we activated AGRP neurons while perturbing their output at downstream circuit nodes. We focused on three brain areas that receive AGRP neuron axonal inputs and also have an established regulatory role in feeding behaviour: pro-opiomelanocortin (*Pomc*)-expressing neurons in the ARC^{4,9,10}, parabrachial nucleus (PBN) neurons in the hindbrain^{11,12}, and paraventricular hypothalamus (PVH) neurons^{13–15} (Supplementary Fig. 1). It is unclear which of these pathways mediate the marked short-term (<1 h) feeding behaviours evoked by AGRP neuron activation.

We used optogenetic and pharmacogenetic tools to examine these complex, molecularly defined feeding circuits in genetically modified mice. Our approach (Supplementary Fig. 2) involves first determining connectivity and synaptic properties between molecularly defined neurons. We then investigated functional consequences of mapped synaptic connections by perturbing electrical activity in pre- and post-synaptic cell types both independently and in concert while monitoring the behavioural response. This is an experimental approach derived from the logic of epistasis analysis¹⁶, used to order mutations into functional pathways or to establish the functional significance of protein interaction networks¹⁷. Here, we have extended this approach

to neuronal activity perturbations in order to establish the functional significance of molecularly defined neural circuits in behaving mice. Our results constrain existing circuit models and also reveal new circuit connections that are directly involved in mediating acute AGRP neuron-evoked feeding behaviour.

Intra-ARC connectivity and function

We first considered the interaction of AGRP neurons in a local circuit with intermingled POMC neurons (Fig. 1a), a population that suppresses food intake^{4,9}. In light of functional opposition between these neurons, their interaction has been implicated as a critical control point for feeding behaviour^{18,19}.

Channelrhodopsin-assisted circuit mapping²⁰ was used to precisely test the functional connectivity matrix of four possible direct synaptic interactions between these two populations (Supplementary Fig. 3a). $ARC^{AGRP} \rightarrow ARC^{POMC}$ connections were probed in *Agrp-Cre; Pomc-TopazFP* bi-transgenic mice in which AGRP neurons were rendered photo-excitable with channelrhodopsin-2 (ref. 21) fused to tdTomato (ChR2:tdTomato), using a Cre recombinase (Cre)-dependent recombinant adeno-associated virus (rAAV)²² (Fig. 1a, b). In brain slices, synaptic currents were recorded from POMC neurons while photo-stimulating AGRP neurons and their axons, which evoked reliable monosynaptic responses in all POMC neurons tested ($n = 19$) (Fig. 1b, c and Supplementary Fig. 4a–e). $ARC^{AGRP} \rightarrow ARC^{POMC}$ synapses showed paired-pulse depression, indicating high release probability (Supplementary Fig. 4f, g). Selective activation of $ARC^{AGRP} \rightarrow ARC^{POMC}$ synaptic connections strongly inhibited POMC neuron activity (Fig. 1d). This circuit connection was blocked by picrotoxin (PTX) (Fig. 1b), a GABA_A (gamma-aminobutyric acid) receptor antagonist, showing that synaptic transmission and POMC neuron suppression required GABA and further indicating that other neuromodulatory substances released under these conditions are not sufficient to silence POMC neurons (Supplementary Fig. 3c). In contrast, no synaptic responses were observed for $ARC^{AGRP} \rightarrow ARC^{AGRP}$, $ARC^{POMC} \rightarrow ARC^{AGRP}$, or $ARC^{POMC} \rightarrow ARC^{POMC}$ (but see ref. 23) connections tested with other transgenic mouse line combinations (Fig. 1c, Supplementary Fig. 3 and Methods). Axon-attached electrophysiological recordings confirmed

¹Janelia Farm Research Campus, Howard Hughes Medical Institute, 19700 Helix Drive Ashburn, Virginia 20147, USA.

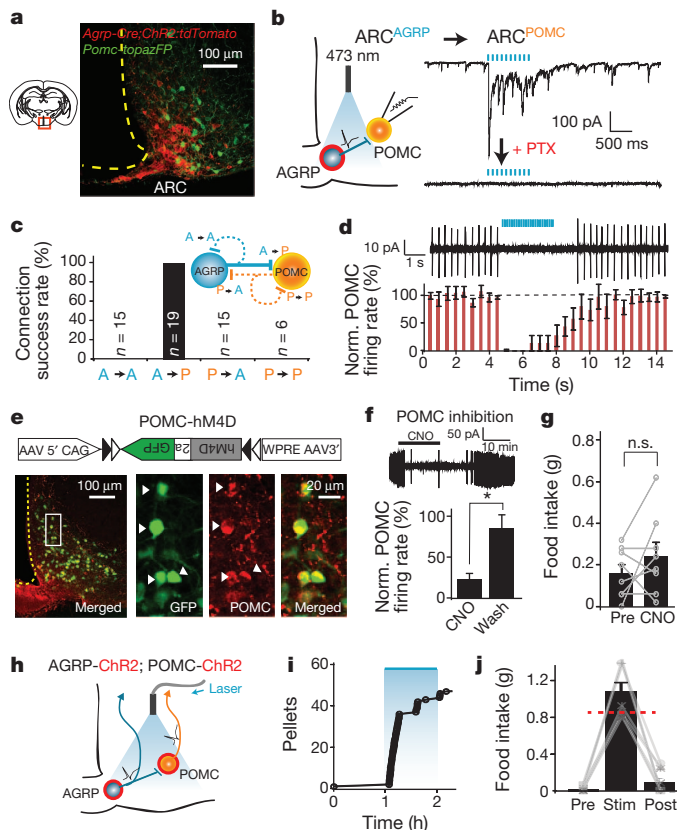


Figure 1 | $ARC^{AGRP} \rightarrow ARC^{POMC}$ is not required for evoked feeding. **a**, ARC in *Agrp-Cre;Pomc-TopazFP* mice expressing ChR2:tdTomato in AGRP neurons. **b**, Left, scheme for testing $ARC^{AGRP} \rightarrow ARC^{POMC}$ synaptic connections. Red, ChR2:tdTomato. Right, $ARC^{AGRP} \rightarrow ARC^{POMC}$ synaptic currents. Blue, light pulses. **c**, Synaptic connectivity between AGRP and POMC neurons. Dashed lines, not detected. **d**, Cell-attached recording (top) and normalized firing rate (bottom, $n = 5$) from $ARC^{AGRP} \rightarrow ARC^{POMC}$ photostimulation in brain slices (with NPY1R, NPY5R and GABA_B receptor antagonists). **e**, POMC neuron expression of hM4D and GFP from Cre-dependent rAAV. **f**, Top, hM4D agonism with CNO (10 μ M). Bottom, firing rate normalized to baseline (paired t -test, $n = 4$). **g**, Intraperitoneal CNO (5 mg kg⁻¹) did not increase food intake (1 h) in POMC-hM4D mice (paired t -test, $P = 0.32$, $n = 8$). **h–j**, Occlusion of $ARC^{AGRP} \rightarrow ARC^{POMC}$ inhibition by optical co-stimulation of AGRP and POMC neurons (**h**) did not impair the feeding response (**i**, **j**) ($n = 9$). Pre, Stim, Post: before, during (blue), after photostimulation (1 h each). Red line, AGRP neuron-evoked food intake from ref. 4. Values are means \pm s.e.m. n.s., not significant, $*P < 0.05$.

that connection probability differences were not a consequence of axonal photo-excitability discrepancies (Supplementary Fig. 5). Thus, comprehensive dissection of connectivity in this molecularly defined circuit reveals striking synaptic specificity.

To test acute behavioural consequences of the $ARC^{AGRP} \rightarrow ARC^{POMC}$ inhibitory connection, we first investigated the sufficiency of POMC neuron inhibition to influence feeding. Cre-dependent rAAV²² was used to target the pharmacogenetic activity silencer hM4D²⁴ to POMC neurons in *Pomc-Cre* mice (POMC-hM4D mice; Fig. 1e). An hM4D agonist, clozapine-N-oxide (CNO, 10 μ M), suppressed POMC neuron activity (Fig. 1f). However, intraperitoneal administration of CNO (5 mg kg⁻¹) to POMC-hM4D mice did not significantly alter food intake over 1 h (Fig. 1g and Supplementary Fig. 6a, b). Chronic POMC neuron suppression (24 h) did increase food intake, and this was dependent on efficient, bilateral transduction of POMC neurons (Supplementary Fig. 6c–e). Thus, POMC neuron suppression does not acutely activate feeding behaviour but does influence long-term food intake.

We also tested the possibility that POMC neuron inhibition might be required for AGRP-neuron-evoked eating. A traditional approach,

such as targeted injection of a GABA_A receptor antagonist while stimulating AGRP neurons, is not cell-type-specific and would release from inhibition all neurons at the injection site (Supplementary Fig. 2c). To overcome this lack of specificity, we developed a cell-type-specific occlusion test to evaluate the behavioural necessity of inhibitory connections between molecularly defined neuron populations by simultaneously co-activating AGRP and POMC neurons to surmount $ARC^{AGRP} \rightarrow ARC^{POMC}$ inhibition (Fig. 1h and Supplementary Fig. 2c).

AGRP and POMC neurons in *Agrp-Cre;Pomc-Cre* bi-transgenic mice were co-transduced with ChR2:tdtomato. We first determined that $ARC^{AGRP} \rightarrow ARC^{POMC}$ inhibitory input could be overcome by ChR2-mediated excitation in POMC neurons (Supplementary Fig. 7). In behavioural experiments, co-activation of AGRP and POMC neurons showed robust AGRP neuron-evoked feeding and rapid latency to eat, even with high ChR2 transduction efficiency in POMC neurons (Fig. 1i, j and Supplementary Fig. 7g, h). Therefore, suppression of POMC neuron activity by AGRP neurons is not required for acute feeding.

Long-range AGRP neural circuit function

We next considered long-range synaptic targets of AGRP neurons⁸. We focused on projections to the PVH in the hypothalamus and to the PBN in the hindbrain because PVH lesions lead to hyperphagia and obesity¹³, whereas pharmacological inhibition in the PBN promotes feeding²⁵ and rescues aphagia induced by AGRP neuron ablation¹¹. However, the relative role for these two projections in AGRP-neuron-evoked feeding behaviour is uncertain. To directly examine these circuits, we used rapid cell-type-specific activation of AGRP axons in the PVH and the PBN.

ChR2-expressing AGRP axons were stimulated with light pulses from an optical fibre placed above either the PVH or the PBN (Fig. 2a). Photoactivation of $ARC^{AGRP} \rightarrow PVH$ axons elicited food intake with similar magnitude as stimulating AGRP-expressing somata in the ARC (Fig. 2b). Conversely, $ARC^{AGRP} \rightarrow PBN$ axon activation did not significantly increase feeding (Fig. 2c, d). Moreover, for mice in which optical fibres were placed above both the PVH and the PBN, robust feeding was only evoked with illumination of the PVH, and there was no significant correlation ($r = 0.17$, $P = 0.68$) for food intake

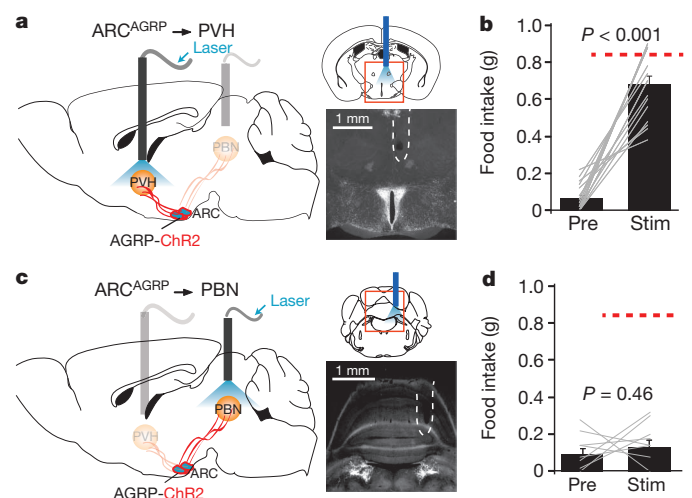


Figure 2 | Feeding is evoked by AGRP axon stimulation in the PVH but not the PBN. **a**, Scheme for AGRP axon photostimulation in the PVH and ChR2:tdTomato distribution of bilaterally transduced AGRP axons in the PVH. Dashed line, optical fibre position. **b**, Food intake before and during photostimulation (1 h each) of AGRP axons in the PVH ($n = 16$ mice, 8 bilateral and 8 unilateral ChR2 transduction). Red line, somatic AGRP-neuron-evoked food intake (ref. 4). **c**, **d**, Photostimulation of AGRP axons with a second optical fibre implanted over the PBN (the bilaterally transduced mice used in **b**). Paired t -test. Values are means \pm s.e.m.

during photostimulation of these two brain areas (Supplementary Fig. 8a). We also observed that the proportion of AGRP-containing axons co-expressing ChR2:tdtomato diverged between the PBN and the PVH, with several rAAV-transduced mice showing ChR2-penetrance biased to the PVH (Supplementary Fig. 8b and see Methods). The dissociation of ChR2-penetrance in AGRP axons between these two brain regions was also apparent from linear regression analysis ($r = 0.83$, $P = 0.006$; y -intercept > 0 , $P = 0.01$, Supplementary Fig. 8b), indicating that not all PVH-projecting AGRP neurons also project to the PBN. Moreover, $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ activation evoked feeding in mice with modest ChR2-penetrance to AGRP axons, whereas $\text{ARC}^{\text{AGRP}} \rightarrow \text{PBN}$ -evoked feeding was not clearly evident even with high ChR2-penetrance (Supplementary Fig. 8c). Therefore, activation of axons in and around the PVH is substantially more effective than those in the PBN for acute, AGRP-neuron-mediated induction of feeding behaviour. In light of these findings, we investigated $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ circuit properties and their relationship to food seeking and consumption behaviour.

$\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ synapse properties

Feeding evoked by activation of AGRP axons in the PVH does not conclusively demonstrate that the PVH is the downstream target. This behavioural effect may also involve activation of AGRP neuron axons-of-passage and antidromic action potentials that could activate AGRP neuron projections to other brain regions. To address this, we investigated AGRP neural circuit connections in the PVH.

We first determined whether AGRP axons in the PVH made synapses that could influence PVH neuronal function. Immunohistochemical analysis revealed AGRP-containing synaptic release sites in the PVH (Supplementary Fig. 9a). Functional synaptic connectivity was measured with channelrhodopsin-assisted circuit mapping²⁰ in the PVH. Photostimulation of AGRP axons evoked synaptic currents in PVH neurons (29/61 cells, 48%). These responses were blocked by PTX (9/9 cells), indicating that they were mediated by GABA (Fig. 3a).

GABA-releasing synapses can have specializations that are essential to their circuit function^{26,27}. Optogenetic methods were used to characterize $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ synaptic connections. The most prominent synaptic characteristic was a barrage of delayed asynchronous inhibitory postsynaptic currents (IPSCs) that continued for up to 1 s following AGRP axon stimulation by a single 1 ms light pulse (Fig. 3b). This property was also observed in other AGRP neuron projection targets (Fig. 1b and Supplementary Fig. 10). These experiments were performed with glutamate receptor antagonists to minimise possible network interactions, and multiple lines of evidence demonstrate that asynchronous release is not a result of photostimulation artefacts (Supplementary Fig. 9). Moreover, axon-attached recordings show that each photostimulus elicits only one axonal action potential (Supplementary Fig. 5).

Asynchronous release increased during repetitive stimulation of $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ projections and synapses quickly developed a slow (DC) component²⁷ that did not return to baseline between the optical stimuli (Fig. 3c, d). Moreover, GABA continued to be released following the stimulus train, and the final IPSC decayed nearly 30-fold more slowly than the underlying $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ quantal IPSCs (Fig. 3e and Methods), which corresponded to a large increase in charge transfer to the postsynaptic neuron after the evoked synchronous IPSC of the last photostimulus (Fig. 3f). Both aspects of asynchronous release were significantly reduced with the membrane-permeable calcium buffer, ethylene glycol-bis(β -aminoethyl)-N,N,N',N'-tetraacetoxymethyl ester (EGTA-AM, 100 μM), indicating that accumulation of free calcium in the synaptic terminal contributes to this property (Fig. 3e, f).

Prolonged asynchronous GABA release at $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ synapses reduced both spontaneous (Fig. 3g) and evoked (Supplementary Fig. 11) PVH excitability for hundreds of milliseconds following IPSCs, a timescale often associated with neuromodulation, but

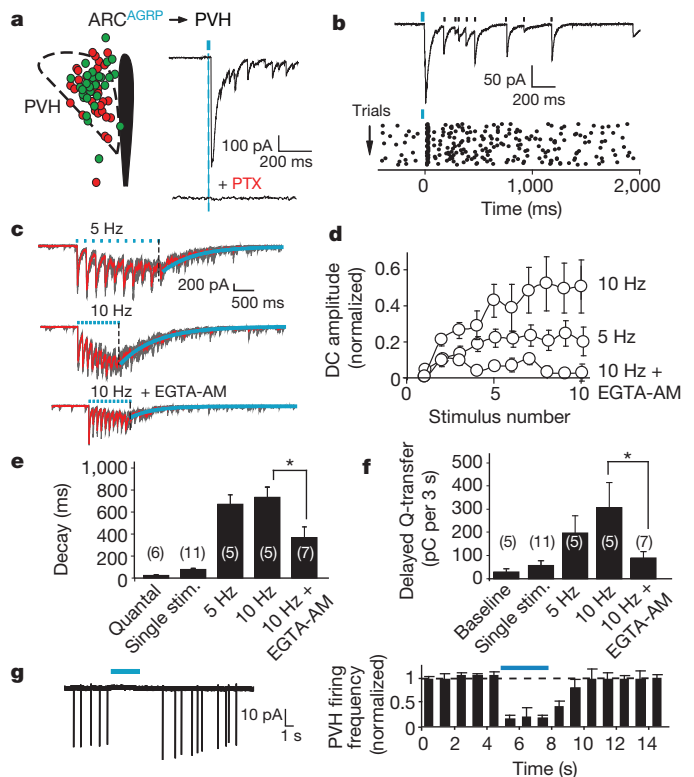


Figure 3 | Prolonged inhibition by $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ synapses.

a, Distribution of $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ postsynaptic neurons. Green, synaptic current detected; red, not detected. Blue tick, light pulse. **b**, Representative $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ IPSC (above); raster plot of IPSCs (below). **c**, $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ repetitive photostimulation. Grey, trials; red, average; blue, post-stimulus exponential fit. **d**, Normalized DC amplitude (see Methods). **e**, **f**, Post-stimulus delayed release: decay time constant (**e**) and cumulative charge transfer (**f**) (sample sizes in parentheses, unpaired t -test). **g**, Left, PVH neuron silenced by AGRP axon photostimulation (10 Hz, with NPY1R, NPY5R and GABA_B receptor antagonists). Right, normalized average firing rates ($n = 9$). Values are means \pm s.e.m. * $P < 0.05$.

mediated here solely by action at ionotropic GABA receptors (Fig. 3g and Supplementary Fig. 11d). This property confers strong inhibition to these synapses, which can influence postsynaptic activity for hundreds of times longer than each AGRP neuron action potential.

Sufficiency and necessity of $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$

Multiple observations suggest that AGRP neurons increase feeding, in part, by inhibiting neurons in the PVH: $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ synaptic connections are specialized to strongly inhibit postsynaptic targets (Fig. 3), this projection is associated with AGRP neuron-evoked eating (Fig. 2), and we have found that food deprivation increases inhibitory synaptic drive onto PVH neurons (Supplementary Fig. 12). Also, GABA_A receptor agonist injections around the ventral thalamus and medial hypothalamus lead to eating²⁸. Therefore, we investigated the causal relationship between feeding behaviour and selective PVH neuron inhibition.

The gene *Sim1* is expressed with regional selectivity in nearly all PVH neurons²⁹. We used *Sim1-Cre* transgenic mice to target the neuronal silencer, hM4D, to neurons in the PVH (SIM1-hM4D mice, Fig. 4a). CNO significantly suppressed SIM1 neuron electrical activity (Supplementary Fig. 13a, b). In SIM1-hM4D mice, CNO increased food intake (Fig. 4b, c and Supplementary Fig. 13c), similar to AGRP neuron photostimulation⁴ and in marked contrast to POMC neuron silencing (Fig. 1g).

Because hM4D acts through a G_i-coupled signalling pathway, the effect on feeding might not be a direct result of neuronal inhibition. Therefore, we used an alternative pharmacologically selective neuronal

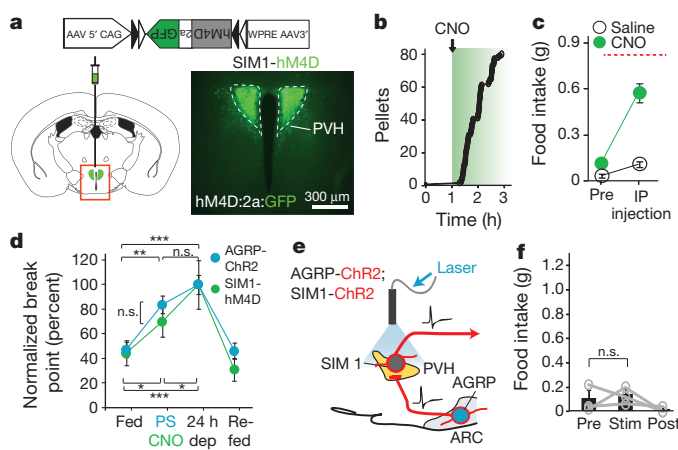


Figure 4 | PVH neuron inhibition recapitulates feeding from AGRP neuron activation. **a**, Cre-dependent rAAV expression of hM4D and GFP in PVH SIM1 neurons. **b**, Food intake before and after intraperitoneal CNO (5 mg kg^{-1}) in a SIM1-hM4D mouse. **c**, Food intake before and after CNO ($n = 19$) or saline ($n = 9$). IP, intraperitoneal. Red line, AGRP neuron-evoked feeding (ref. 4). **d**, Normalized break point for AGRP-ChR2 ($n = 6$) or SIM1-hM4D ($n = 6$) mice (two-way ANOVA with 1-factor repetition, condition: $F_{2,20} = 25.5$, $P < 0.001$; AGRP/SIM1: $F_{1,20} = 0.171$, $P = 0.69$; interaction: $F_{2,20} = 0.45$, $P = 0.64$; Holm-Sidak correction for multiple comparisons). Normalized to food deprivation (dep) break point. **e**, **f**, Scheme for co-photostimulation of AGRP axons and SIM1 neurons over the PVH (**e**), and evoked food intake (**f**) before, during and after (1 h each) co-stimulation (paired t -test, $n = 3$). Values are means \pm s.e.m. n.s., not significant; $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

silencer system which can effectively mimic prolonged ionotropic GABA receptor activation by directly increasing chloride conductance³⁰. Silencing SIM1 neurons with PSAM^{L141F}-GlyR chimaeric ion channels and a cognate selective ligand (PSEM³⁰⁸) also rapidly evoked feeding (Supplementary Fig. 13d, e). Both of these experiments are consistent with acute PVH inhibition as an important contributor to AGRP neuron-evoked feeding behaviour.

Deprivation-induced hunger and AGRP neuron activation are both associated with an increased willingness to work for food^{5,31}. To explore whether AGRP neural circuits could be used to identify new motivationally-sensitive brain regions, we compared the motivational shift induced by AGRP neuron activation, SIM1 neuron silencing and food deprivation in an instrumental lever press task. We used progressive ratio (PR) food pellet reinforcement, where the highest reinforcement schedule attained is termed the break point, a measure of motivation³¹. Well-fed AGRP-ChR2 and SIM1-hM4D mice both increased break point during either AGRP neuron activation or SIM1 neuron silencing (Fig. 4d), but not in control mice (Supplementary Fig. 13f). Together, these results show that both food seeking and food consumption responses to AGRP neuron activation are replicated by suppressing the electrical activity of SIM1 neurons. Moreover, circuit mapping approaches allowed us to identify the PVH, which has not been previously implicated in instrumental responses for reinforcement, as a motivationally important brain region.

We also tested the necessity of this circuit connection for elevated food intake using co-activation of SIM1 neurons to selectively overcome ARC^{AGRP}→PVH inhibition (Fig. 4e). In PVH-containing brain slices from *Agrp-Cre;Sim1-Cre* bi-transgenic mice where both neuron populations were virally transduced to express ChR2, we determined that co-stimulation of SIM1 neurons along with AGRP axons overcame the inhibitory response from ARC^{AGRP}→PVH projections (Supplementary Fig. 13h). *In vivo*, co-stimulation of ARC^{AGRP}→PVH projections and SIM1 neurons completely suppressed AGRP neuron-evoked eating (Fig. 4f). Post hoc analysis showed that AGRP neurons were still activated under these conditions ($\text{Fos}^+/\text{ChR2}^+$: $77 \pm 2\%$, $n = 3$). Thus, PVH inhibition is both necessary and sufficient for acute feeding evoked by ARC^{AGRP}→PVH axon projections.

ARC^{AGRP}→PVH^{OXT} connectivity and function

The PVH is a heterogeneous brain structure with different cell types^{32,33}. Within the PVH, oxytocin (OXT) neuron loss-of-function is implicated in Prader-Willi syndrome³⁴ and also as a consequence of *SIM1* mutations^{29,35,36}, both of which are associated with overeating and obesity in people. To refine ARC^{AGRP}→PVH connectivity, we investigated OXT neurons as potential postsynaptic targets of this inhibitory projection.

Functional circuit mapping experiments showed that OXT neurons, a small subpopulation of PVH neurons³³ identified in *Agrp-Cre;Oxytocin-Gfp* bi-transgenic mice, showed substantially higher connection probability with AGRP neurons (12/17 cells, 71%) than for neighbouring unlabelled neurons (6/18 cells, 33%; Supplementary Fig. 14a–c). Despite its prominence, this ARC^{AGRP}→PVH^{OXT} connection has not been previously described, in part due to the difficulty of conclusively demonstrating cell type-specific synaptic connections without optogenetic electrophysiological methods. In addition, the ARC^{AGRP}→PVH^{OXT} circuit was strongly inhibitory (Supplementary Fig. 14d), therefore AGRP neurons may activate feeding through inhibition of OXT neurons.

We tested the behavioural necessity of this circuit by occlusion of synaptic inhibition at this molecularly defined circuit connection with co-activation of OXT neurons and AGRP neuron projections. OXT neurons were rendered photo-excitable by transduction with rAAV expressing ChR2:tdTomato from an oxytocin promoter fragment^{37,38} (Supplementary Fig. 14e–h), and *in vivo* photostimulation increased Fos expression selectively in OXT neurons (Supplementary Fig. 14i, j). Photostimulation over the PVH containing both ChR2-expressing OXT neurons and AGRP axons strongly suppressed evoked feeding (Fig. 5a, b).

To confirm that suppression of feeding was dependent on ARC^{AGRP}→PVH^{OXT} circuit connectivity and not an autonomous OXT neuron pathway, we used a cell-type-specific circuit disconnection strategy. Because ARC^{AGRP}→PVH projections are lateralised (~3:1 ipsilateral:contralateral projection bias), we stimulated unilaterally transduced OXT neurons either ipsilateral or contralateral to the side of unilateral AGRP neuron transduction (Fig. 5c). Ipsilateral OXT neuron stimulation significantly suppressed ARC^{AGRP}→PVH stimulation-induced feeding and contralateral OXT neuron stimulation

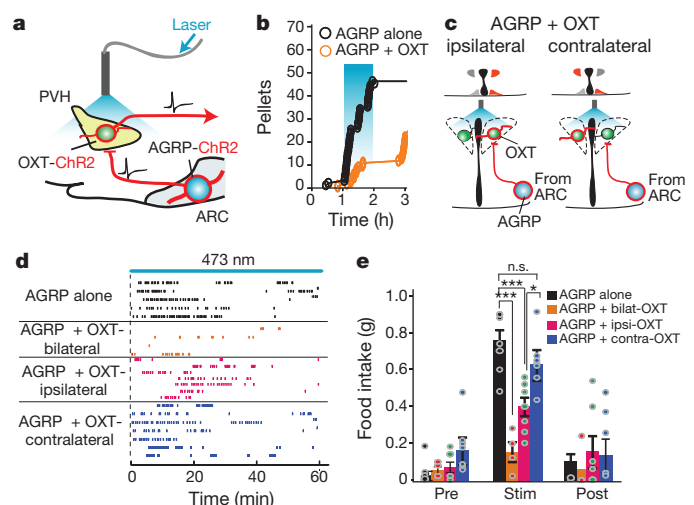


Figure 5 | ARC^{AGRP}→PVH^{OXT} circuit contributes to evoked food intake. **a**, **b**, Co-photostimulation scheme for OXT neurons and AGRP axons over the PVH (**a**) and evoked food intake (**b**). Blue shading, photostimulation. **c**, Disconnection scheme for ARC^{AGRP}→PVH^{OXT}. **d**, Raster plot showing food pellets taken; rows, separate mice. **e**, Food intake before, during and after photostimulation (1 h each). Within photostimulation condition: ANOVA, $F_{3,22} = 18.2$, $P < 0.001$; Holm-Sidak correction for multiple comparisons. Values are means \pm s.e.m. n.s., not significant; $*P < 0.05$, $***P < 0.001$.

did not (Fig. 5d, e and Supplementary Fig. 14k). Occlusion of the ipsilateral $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}^{\text{OXT}}$ circuit is probably an underestimate because $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ projections are not strictly ipsilateral. Notably, bilateral stimulation of OXT neurons alone in food-deprived mice during re-feeding, which is mediated by multiple cell types and circuits, did not significantly reduce consumption (Supplementary Fig. 14l), thus the anorexigenic effect of OXT neuron activation is most apparent during selective activation of AGRP axons in the PVH. All together, these experiments identify the molecularly defined circuit projection $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}^{\text{OXT}}$ as a synaptic pathway for AGRP neurons in the PVH and demonstrate an important role for this connection in AGRP-neuron-evoked eating.

Pharmacological analysis of evoked feeding

A remaining question concerns the relative behavioural contribution of GABA and neuropeptides released by AGRP neurons. We previously showed that the neuromodulator AGRP was not necessary for acute feeding behaviour⁴, but a role for co-released NPY is possible because this peptide potentially activates feeding when injected to the PVH³⁹. Initial circuit characterization experiments using *Npy*^{-/-} mice to address the necessity of NPY release showed marked strengthening of GABA signalling in $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ circuitry (connectivity rate, *AgRP-Cre*: 48%, *AgRP-Cre;Npy*^{-/-}: 69%; paired-pulse ratio, *AgRP-Cre*: 0.81 ± 0.08 , *n* = 24; *AgRP-Cre;Npy*^{-/-}: 0.46 ± 0.09 , *n* = 28; *P* = 0.01, unpaired *t*-test), likely due to substantial developmental compensation in these circuits. Therefore, to probe the contribution of these transmitters to $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ neuron-evoked food intake while minimizing compensatory effects, we pharmacologically blocked GABA_A receptors and NPY1R in the PVH during $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ axon activation. Blockade of either receptor strongly inhibited evoked food intake during $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ photostimulation (Fig. 6a, b), indicating that concerted signalling through both receptor types was essential for the behavioural effects of AGRP axon activation in the PVH.

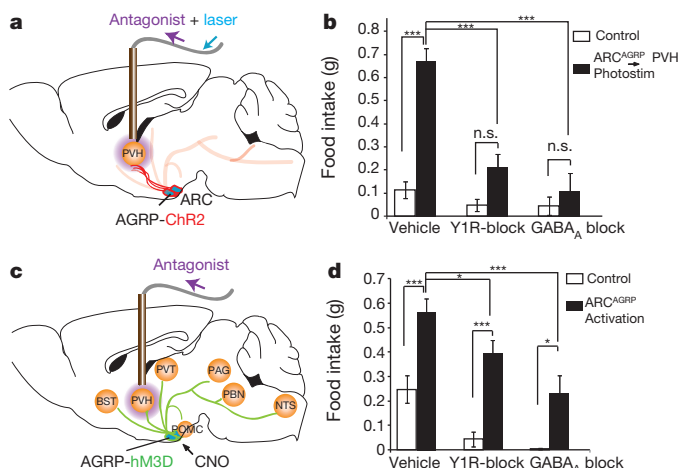


Figure 6 | Pharmacological dissection of AGRP neuron-evoked feeding. **a**, Scheme depicting infusion of antagonists (purple) for GABA_A receptors or NPY1R into the PVH followed by $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ axon photostimulation. **b**, Food intake during photostimulation (1 h) in ChR2-expressing mice (black, *n* = 5) and ChR2-negative mice (open bars, *n* = 5). Two-way ANOVA with 1-factor repetition; $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ activation: $F_{1,16} = 24.4$, *P* = 0.001; antagonist: $F_{2,16} = 13.0$, *P* < 0.001; interaction: $F_{2,16} = 8.6$, *P* = 0.003. **c**, As in **a** except with somatic AGRP neuron activation using hM3D and CNO (0.3 mg kg⁻¹). **d**, Food intake (1 h) after intraperitoneal CNO injection in hM3D-expressing (black, *n* = 8) and non-transduced control mice (open bars, *n* = 5). Two-way ANOVA with 1-factor repetition; neuron activation: $F_{1,20} = 22.8$, *P* < 0.001; antagonist: $F_{2,20} = 14.5$, *P* < 0.001; interaction: $F_{2,20} = 0.37$, *P* = 0.694. Holm–Sidak correction for multiple comparisons. Values are means \pm s.e.m. n.s., not significant; **P* < 0.05, ****P* < 0.001.

Robust pharmacological suppression of feeding in the $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ circuit allowed us to use pharmacology to examine the overall necessity of this projection pathway in the context of somatic AGRP neuron-evoked eating, which is expected to activate all AGRP neuron circuit projections. For this, we activated AGRP neurons, using the pharmacogenetic activator hM3D with its ligand CNO^{5,24}, and blocked either GABA or NPY signalling in the PVH (Fig. 6c). GABA_A or NPY1R antagonists in the PVH significantly reduced food intake during AGRP neuron activation (Fig. 6d). These results indicate that, even with brain-wide activation of AGRP neural circuits, signalling in the PVH is critical and involves both GABA_A and NPY receptors. Strikingly, though, in these experiments, and in contrast to $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ projection activation, feeding remained significantly above baseline (Fig. 6d). Thus, although our results show that AGRP neuron-evoked eating is mediated, in part, by the PVH, there are additional behaviourally important circuits awaiting further investigation.

Discussion

Hunger is mediated by neural circuits that integrate visceral signals of energetic state and consequently regulate physiology and behaviour. Applying optogenetic circuit mapping, we determined functional connectivity from starvation-sensitive AGRP neurons to synaptic targets with cell type-specific precision and then evaluated the functional significance of these connections for feeding behaviour by manipulating the circuit nodes independently or together in behaving mice (Supplementary Fig. 15a). Based on these experiments and the results of others, three distinct functions can be assigned to anatomically separate projection fields of AGRP neurons (Supplementary Fig. 15b). 1) We show directly that AGRP neurons strongly inhibit POMC neurons, which do not acutely regulate feeding behaviour. However, cell type-specific activation⁴ and silencing experiments indicate that they likely regulate longer-term feeding responses. 2) We find prolonged inhibition of PVH neurons by synapses from AGRP axons. Suppression of PVH neurons is sufficient for acute AGRP neuron-evoked eating, which also includes the motivational consequences of AGRP neuron activation. In addition, the $\text{ARC}^{\text{AGRP}} \rightarrow \text{PVH}$ circuit projection is necessary for a significant portion of AGRP neuron-evoked eating; however other brain regions likely contribute. 3) Finally, AGRP neuron projections targeting the parabrachial nucleus (PBN) in the hindbrain do not directly activate feeding, but instead they restrain visceral malaise that results from AGRP neuron ablation¹¹. Future experiments could investigate the possibility that this is due to separate AGRP neuron subpopulations in the ARC with different axonal projection patterns. In any event, multiple mechanisms involved in hunger are dissociated into distinct behavioural modules by anatomically separate AGRP neuron axonal projections.

These experiments also support an important link between forebrain and hindbrain control of feeding behaviour. In the forebrain, we have identified OXT neurons as a key target of AGRP neurons for controlling acute feeding behaviour. OXT potently suppresses feeding when delivered to the brain but not the periphery⁴⁰, thus non-neuroendocrine OXT neurons, which project to the hindbrain and spinal cord^{33,41}, likely mediate these anorexigenic effects. Oxytocin signalling enhances hindbrain responses to circulating satiety signals⁴², and genetic disruption of synaptic release from OXT neurons⁴³ or ablation of OXT receptor-expressing hindbrain neurons both lead to overeating⁴⁴. Loss of OXT neurons is also associated with both Prader–Willi syndrome and *SIM1* mutations, each of which lead to insatiable hunger in people, presumably due to a disrupted satiety response^{29,34–36}. These experiments define a circuit for voracious AGRP neuron-evoked eating and link these neurons that regulate energy homeostasis to hindbrain-projecting neurons involved with human genetic conditions that lead to remarkably similar characteristics to the evoked behaviour: profound hunger, motivation for food, and

resultant obesity. This mechanistic relationship implies that AGRP neuron activation may be a behavioural model that could be used to explore therapeutic approaches to overeating observed in these conditions. Further investigation into the “hunger modules” described here using cell type-specific mapping and manipulation techniques will permit elaboration of this circuit framework for feeding regulation and provide insight into this behaviour under healthy conditions and in association with overeating disorders.

METHODS SUMMARY

All experimental protocols were conducted according to US National Institutes of Health guidelines for animal research and were approved by the Institutional Animal Care and Use Committee at Janelia Farm Research Campus.

Optogenetic experiments. Light was delivered to the brain through an optical fibre positioned ~0.8 mm from the targeted region. The light power exiting the fibre (10–15 mW) was estimated to correspond to $>2.0 \text{ mW mm}^{-2}$ at the ARC, PVH or PBN. The photostimulation protocol was 10 ms pulses, 20 pulses for 1 s, repeated every 4 s for 1 h.

Full Methods and any associated references are available in the online version of the paper.

Received 26 September 2011; accepted 29 May 2012.

Published online 11 July 2012.

- Cowley, M. A. *et al.* The distribution and mechanism of action of ghrelin in the CNS demonstrates a novel hypothalamic circuit regulating energy homeostasis. *Neuron* **37**, 649–661 (2003).
- Andrews, Z. B. *et al.* UCP2 mediates ghrelin's action on NPY/AgRP neurons by lowering free radicals. *Nature* **454**, 846–851 (2008).
- Yang, Y., Atasoy, D., Su, H. H. & Sternson, S. M. Hunger states switch a flip-flop memory circuit via a synaptic AMPK-dependent positive feedback loop. *Cell* **146**, 992–1003 (2011).
- Aponte, Y., Atasoy, D. & Sternson, S. M. AGRP neurons are sufficient to orchestrate feeding behavior rapidly and without training. *Nature Neurosci.* **14**, 351–355 (2011).
- Krashes, M. J. *et al.* Rapid, reversible activation of AgRP neurons drives feeding behavior in mice. *J. Clin. Invest.* **121**, 1424–1428 (2011).
- Luquet, S., Perez, F. A., Hnasko, T. S. & Palmiter, R. D. NPY/AgRP neurons are essential for feeding in adult mice but can be ablated in neonates. *Science* **310**, 683–685 (2005).
- Cannon, W. B. *The Wisdom of the Body* 2nd edn (W.W. Norton & Co., 1939).
- Broberger, C., Johansen, J., Johansson, C., Schalling, M. & Hokfelt, T. The neuropeptide Y/agouti gene-related protein (AGRP) brain circuitry in normal, anorectic, and monosodium glutamate-treated mice. *Proc. Natl Acad. Sci. USA* **95**, 15043–15048 (1998).
- Gropp, E. *et al.* Agouti-related peptide-expressing neurons are mandatory for feeding. *Nature Neurosci.* **8**, 1289–1291 (2005).
- Yaswen, L., Diehl, N., Brennan, M. B. & Hochgeschwender, U. Obesity in the mouse model of pro-opiomelanocortin deficiency responds to peripheral melanocortin. *Nature Med.* **5**, 1066–1070 (1999).
- Wu, Q., Boyle, M. P. & Palmiter, R. D. Loss of GABAergic signaling by AgRP neurons to the parabrachial nucleus leads to starvation. *Cell* **137**, 1225–1234 (2009).
- Wu, Q., Clark, M. S. & Palmiter, R. D. Deciphering a neuronal circuit that mediates appetite. *Nature* **483**, 594–597 (2012).
- Leibowitz, S. F., Hammer, N. J. & Chang, K. Hypothalamic paraventricular nucleus lesions produce overeating and obesity in the rat. *Physiol. Behav.* **27**, 1031–1040 (1981).
- Balthasar, N. *et al.* Divergence of melanocortin pathways in the control of food intake and energy expenditure. *Cell* **123**, 493–505 (2005).
- Xi, D., Gandhi, N., Lai, M. & Kublaoui, B. M. Ablation of Sim1 neurons causes obesity through hyperphagia and reduced energy expenditure. *PLoS ONE* **7**, e36453 (2012).
- Avery, L. & Wasserman, S. Ordering gene function: the interpretation of epistasis in regulatory hierarchies. *Trends Genet.* **8**, 312–316 (1992).
- Collins, S. R. *et al.* Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* **446**, 806–810 (2007).
- Cowley, M. A. *et al.* Leptin activates anorexigenic POMC neurons through a neural network in the arcuate nucleus. *Nature* **411**, 480–484 (2001).
- Tong, Q., Ye, C. P., Jones, J. E., Elmquist, J. K. & Lowell, B. B. Synaptic release of GABA by AgRP neurons is required for normal regulation of energy balance. *Nature Neurosci.* **11**, 998–1000 (2008).
- Petreaanu, L., Huber, D., Sobczyk, A. & Svoboda, K. Channelrhodopsin-2-assisted circuit mapping of long-range callosal projections. *Nature Neurosci.* **10**, 663–668 (2007).
- Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neurosci.* **8**, 1263–1268 (2005).
- Atasoy, D., Aponte, Y., Su, H. H. & Sternson, S. M. A FLEX switch targets channelrhodopsin-2 to multiple cell types for imaging and long-range circuit mapping. *J. Neurosci.* **28**, 7025–7030 (2008).
- Dicken, M. S., Tooker, R. E. & Hentges, S. T. Regulation of GABA and glutamate release from proopiomelanocortin neuron terminals in intact hypothalamic networks. *J. Neurosci.* **32**, 4042–4048 (2012).
- Armbruster, B. N., Li, X., Pausch, M. H., Herlitze, S. & Roth, B. L. Evolving the lock to fit the key to create a family of G protein-coupled receptors potentially activated by an inert ligand. *Proc. Natl Acad. Sci. USA* **104**, 5163–5168 (2007).
- Higgs, S. & Cooper, S. J. Hyperphagia induced by direct administration of midazolam into the parabrachial nucleus of the rat. *Eur. J. Pharmacol.* **313**, 1–9 (1996).
- Hefft, S. & Jonas, P. Asynchronous GABA release generates long-lasting inhibition at a hippocampal interneuron-principal neuron synapse. *Nature Neurosci.* **8**, 1319–1328 (2005).
- Best, A. R. & Regehr, W. G. Inhibitory regulation of electrically coupled neurons in the inferior olive is mediated by asynchronous release of GABA. *Neuron* **62**, 555–565 (2009).
- Kelly, J., Rothstein, J. & Grossman, S. P. GABA and hypothalamic feeding systems. I. Topographic analysis of the effects of microinjections of muscimol. *Physiol. Behav.* **23**, 1123–1134 (1979).
- Kublaoui, B. M., Gemelli, T., Tolson, K. P., Wang, Y. & Zinn, A. R. Oxytocin deficiency mediates hyperphagic obesity of *Sim1* haploinsufficient mice. *Mol. Endocrinol.* **22**, 1723–1734 (2008).
- Magnus, C. J. *et al.* Chemical and genetic engineering of selective ion channel-ligand interactions. *Science* **333**, 1292–1296 (2011).
- Hodos, W. Progressive ratio as a measure of reward strength. *Science* **134**, 943–944 (1961).
- Swanson, L. W. & Sawchenko, P. E. Hypothalamic integration: organization of the paraventricular and supraoptic nuclei. *Annu. Rev. Neurosci.* **6**, 269–324 (1983).
- Biag, J. *et al.* Cyto- and chemoarchitecture of the hypothalamic paraventricular nucleus in the C57BL/6J male mouse: a study of immunostaining and multiple fluorescent tract tracing. *J. Comp. Neurol.* **520**, 6–33 (2012).
- Swaab, D. F., Purba, J. S. & Hofman, M. A. Alterations in the hypothalamic paraventricular nucleus and its oxytocin neurons (putative satiety cells) in Prader–Willi syndrome: a study of five cases. *J. Clin. Endocrinol. Metab.* **80**, 573–579 (1995).
- Holder, J. L., Jr, Butte, N. F. & Zinn, A. R. Profound obesity associated with a balanced translocation that disrupts the *SIM1* gene. *Hum. Mol. Genet.* **9**, 101–108 (2000).
- Traurig, M. *et al.* Common variation in *SIM1* is reproducibly associated with BMI in Pima Indians. *Diabetes* **58**, 1682–1689 (2009).
- Fields, R. L., Donzio, T. A., Kawasaki, M. & Gainer, H. Cell-type specific oxytocin gene expression from AAV delivered promoter deletion constructs into the rat supraoptic nucleus *in vivo*. *PLoS ONE* **7**, e32085 (2012).
- Knobloch, H. S. *et al.* Evoked axonal oxytocin release in the central amygdala attenuates fear response. *Neuron* **73**, 553–566 (2012).
- Clark, J. T., Kalra, P. S., Crowley, W. R. & Kalra, S. P. Neuropeptide Y and human pancreatic polypeptide stimulate feeding behavior in rats. *Endocrinology* **115**, 427–429 (1984).
- Arletti, R., Benelli, A. & Bertolini, A. Oxytocin inhibits food and fluid intake in rats. *Physiol. Behav.* **48**, 825–830 (1990).
- Swanson, L. W., Sawchenko, P. E., Wiegand, S. J. & Price, J. L. Separate neurons in the paraventricular nucleus project to the median eminence and to the medulla or spinal cord. *Brain Res.* **198**, 190–195 (1980).
- Blevins, J. E., Eakin, T. J., Murphy, J. A., Schwartz, M. W. & Baskin, D. G. Oxytocin innervation of caudal brainstem nuclei activated by cholecystokinin. *Brain Res.* **993**, 30–41 (2003).
- Zhang, G. *et al.* Neuropeptide exocytosis involving synaptotagmin-4 and oxytocin in hypothalamic programming of body weight and energy balance. *Neuron* **69**, 523–535 (2011).
- Baskin, D. G. *et al.* A new oxytocin-saporin cytotoxin for lesioning oxytocin-receptive neurons in the rat hindbrain. *Endocrinology* **151**, 4207–4213 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research was funded by the Howard Hughes Medical Institute. We thank J. Cox, A. Wardlaw, K. Morris for mouse breeding, genotyping, and viral injection support; S. Michael and A. Hu for histology support; M. Ramirez and B. Zemelman for rAAV production; H. Gainer for discussions about the oxytocin promoter; and E. Boyden for technical assistance.

Author Contributions D.A., J.N.B. and S.M.S. designed the experiments and analysed data. D.A. and J.N.B. performed experiments. H.H.S. performed molecular cloning for viral constructs. S.M.S. and D.A. wrote the manuscript with comments from all of the authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.M.S. (sternsons@janelia.hhmi.org).

METHODS

All experimental protocols were conducted according to US National Institutes of Health guidelines for animal research and were approved by the Institutional Animal Care and Use Committee at Janelia Farm Research Campus.

Mice. Animals were housed on a 12 h light (06:00)/dark (18:00) cycle with *ad libitum* access to water and mouse chow (PicoLab Rodent Diet 20, 5053 tablet, TestDiet), unless otherwise noted. *AgRP-Cre⁴⁵*, *AgRP-IRES-Cre¹⁹*, *Pomc-Cre⁴⁶*, *Sim1-Cre¹⁴*, *Oxytocin-Gfp⁴⁷*, *Pomc-TopazFP⁴⁸*, *NPY-SapphireFP⁴⁸*, *Npy^{-/-49}* mice have been described previously. In most cases, behavioural experiments were with male mice. Females were used for some SIM1 neuron silencing experiments (6/19, Fig. 4). Most experiments used *AgRP-Cre⁴⁵* mice, except experiments in Fig. 2b (subset), Figs 2c, d, 6 and Supplementary Fig. 8, which were performed with *AgRP-IRES-Cre¹⁹* mice.

Viral vectors. rAAV2/1-CAG-FLEX-*rev-ChR2:tdTomato* was described previously²². The rAAV-CAG-FLEX-*rev-hM4D:2a:GFP* vector was prepared by ligating hM4D and EGFP with an intervening DNA fragment for a 2a peptide from Porcine teschovirus (GSGATNFSLKQAGDVEENPGP), which was inserted into the rAAV2-CAG-FLEX backbone in an inverted orientation. rAAV2/1-*hSyn-FLEX-*rev-hM3D:mCherry** was from the UNC viral core facility. For rAAV2-*Oxytocin-ChR2:tdTomato*, we used a mouse oxytocin promoter fragment (−600 to −1, forward: 5′-CAAGCCAGCCTGGTCTACACAGCAGG-3′, reverse: 5′-GGCGATGGTGTCTGCTGAGATCCCG-3′), followed by the Promega chimaeric intron and ChR2:tdTomato, which were ligated into an rAAV2 backbone. Viral vectors were produced by the University of Pennsylvania Gene Therapy Program Vector Core or the Janelia Farm Molecular Biology Core Facility.

Viral injections and fibre placement. Viral injections were performed as described previously²² (P21–P25 for electrophysiological recordings, P40–P50 for behavioural experiments). ARC coordinates: bregma −1.2 mm, midline +0.2 mm; dorsal surface −5.85 mm and −5.75 mm. PVH coordinates: bregma −0.7 mm; midline ±0.3 mm; dorsal surface −4.5 mm and −4.3 mm. For behavioural experiments that required photostimulation, a guide cannula was inserted (ARC, 4.5 mm, 26GA; PVH, 3.5 mm, 26GA). For PVH^{OXT} neuronal activation, cannula placement was through a midline craniotomy. For PBN/PVH photostimulation in the same mouse, adult male *AgRP-IRES-Cre¹⁹* animals were bilaterally transduced in the arcuate nucleus with rAAV2/1-CAG-FLEX-*rev-ChR2:tdTomato* (600 nl). Ferrule-capped fibres (see below) were implanted over the PVH (bregma: −0.7 mm, midline: +0.3 mm; dorsal surface: −4.0 mm) and PBN (bregma: −5.8 mm, midline: +0.9 mm; dorsal surface: −2.75 mm).

Grip cement (DENTSPLY) was used to anchor the guide cannula or ferrule-capped fibres to the skull. When needed, a dummy cannula (33GA, Plastics One) was inserted to keep the fibre guide from getting clogged. Postoperative analgesia was provided (ketoprofen, 5 mg kg^{−1}). After surgery, mice were allowed 14–20 days for recovery and transgene expression.

Pharmacology. Antagonists: GABA_A (picrotoxin, 50 μM, Sigma), GABA_B (saclofen, 50 μM, Tocris), ionotropic glutamate receptors (AP-5, 50 μM; CNQX, 10 μM; Sigma), NPY1R (PD160170, 1 μM, Tocris), NPY2R (BIEE 0246, 0.5 μM, Tocris), and NPY5R (CGP 71683, 10 μM, Tocris), voltage-gated sodium channels (tetrodotoxin, TTX, 1 μM, Sigma). Saclofen was included to prevent potential metabotropic GABA_B receptor-mediated modulation of the postsynaptic neuron. Agonist: CNO (10 μM, BioMol). Pharmacological agents were bath-applied with gravity perfusion.

Electrophysiology and circuit mapping. Experimental techniques were similar to those reported previously²². Detailed conditions for circuit mapping experiments in brain slices are in Supplementary Table 1. Coronal brain slices were prepared in chilled cutting solution containing (in mM): 234 sucrose, 28 NaHCO₃, 7 dextrose, 2.5 KCl, 7 MgCl₂, 0.5 CaCl₂, 1 sodium ascorbate, 3 sodium pyruvate and 1.25 NaH₂PO₄, aerated with 95% O₂/5% CO₂. Slices were transferred to artificial cerebrospinal fluid (aCSF) containing (in mM): 119 NaCl, 25 NaHCO₃, 11 D-glucose, 2.5 KCl, 1.25 MgCl₂, 2 CaCl₂ and 1.25 NaH₂PO₄, aerated with 95% O₂/5% CO₂. Slices were incubated at 34 °C (30 min) and then maintained and recorded from at room temperature (20–24 °C). The intracellular solution for voltage clamp recordings contained (in mM): 125 CsCl, 5 NaCl, 10 HEPES, 0.6 EGTA, 4 Mg-ATP, 0.3 Na₂GTP, 10 lidocaine *N*-ethyl bromide (QX-314), pH 7.35 and 290 mOsm l^{−1}. The holding potential for voltage clamp recordings was −60 mV unless otherwise indicated. The intracellular solution for current clamp recordings contained (in mM): 125 potassium gluconate, 6.7 KCl, 10 HEPES, 1 EGTA, 4 Mg-ATP, 10 sodium phosphocreatine (pH 7.25; 290 mOsm l^{−1}), *E*_{Cl} = −75 mV. In a subset of experiments potassium gluconate-based internal solution was used for voltage clamp recordings. In most intracellular recordings, internal solutions contained GDP-βS (0.5 mM, Sigma).

For brain slice photostimulation, a laser (473 nm) was used to deliver light pulses ranging from 0.1 to 1 mW at the specimen. Laser power was monitored

with a photodiode for each light pulse. Light pulse duration (1 ms) was controlled by a Pockels cell (ConOptics) and a mechanical shutter (Vincent Associates). A focal spot was targeted onto the specimen with two scanning mirrors (Cambridge Technology) through 4× or 63× objectives.

For electrical stimulation, a field electrode was placed within the ARC (for measurements from ARC) or adjacent to the third ventricle to activate the ascending fibre tract (for measurements from the PVH). Half-maximal stimulus strength was used for asynchronous release or paired-pulse ratio measurements.

Loose-seal, cell-attached recordings (seal resistance, 20–70 MΩ, aCSF internal) were made in voltage clamp with holding current maintained at zero. Most neurons fired spontaneously. For measuring ARC^{AGRP}→ARC^{POMC} and ARC^{AGRP}→PVH^{OXT} influence on spontaneous firing rate, the postsynaptic neurons were recorded while ChR2-expressing axons were photostimulated in the absence of any blockers (Supplementary Fig. 3c and Supplementary Fig. 14d) or in the presence of NPY1R, NPY5R and GABA_B receptor blockers (Fig. 1d and Fig. 3g). hM4D-dependent neuronal silencing was tested in POMC and SIM1 neurons in the presence of glutamate and GABA_A receptor blockers.

For axon-attached recordings, an aCSF-filled recording electrode (8–10 MΩ) was used, and ionotropic glutamate and GABA_A receptors were blocked. After a subset of recordings, TTX (1 μM) was used to confirm that signals were due to action potentials. AGRP or POMC axons in the PVH were identified and targeted by tdTomato fluorescence.

For asynchronous release measurements, CNQX, AP5 and saclofen were present. Asynchronous release was also prominent in the absence of saclofen (Fig. 1b and Fig. 3b). For analysis of asynchronous release, traces from 8–10 trials were averaged. The DC component during photostimulation was calculated by measuring the current amplitude 2 ms before each photostimulus and was normalized to the peak amplitude of the first synchronous synaptic response (Fig. 3d). Decay times for the delayed asynchronous component in the train stimulus were calculated by a single exponential fit starting 100 ms following last light pulse (Fig. 3e). Cumulative charge from delayed release was calculated as the total area under the averaged traces 100 ms–3 s following last pulse (Fig. 3f). For baseline charge transfer a 3 s pre-stimulus window was used.

Quantal amplitude measurements (Fig. 3e) were performed under the same conditions as above except that, in the aCSF, 2 mM Ca²⁺ was replaced by 2 mM Sr²⁺. Quantal events were chosen from a window immediately following stimulation until the event frequency dropped to three times above the baseline spontaneous event frequency.

In vivo photostimulation. Components for food consumption monitoring and photostimulation were similar to those reported previously⁴. Light was delivered to the brain through an optical fibre (200 μm diameter core; BFH48-200-Multimode, NA 0.48; Thorlabs), which was implanted through the fibre guide the day before photostimulation. The fibre tip was positioned to a distance of ~0.8 mm from the targeted region. The relationship of light scattering and absorption in the brain as a function of distance has been described previously⁵⁰. Using this relationship, the light power exiting the fibre tip (10–15 mW) was estimated to correspond to >2.0 mW mm^{−2} at the ARC or PVH. For optical delivery of light pulses with millisecond precision to multiple mice, the output from a diode laser (473 nm, Altechna) was split into eight beams using a combination of 50/50 beam splitters and turning mirrors (Thorlabs). The main output beam from the diode laser was controlled using an acousto-optic modulator (AOM) (Quanta Tech, OPTO-ELECTRONIC) to generate light pulses that were launched into separate fibre ports (PAF-X-5 or PAF-X-7, Thorlabs) and their corresponding optical fibres. Using these components, eight mice could be simultaneously photostimulated. For all *in vivo* photostimulation experiments the same pulse protocol was used: 10 ms pulses, 20 pulses for 1 s, repeated every 4 s for 1 h.

For ARC^{AGRP}→PVH^{OXT} photostimulation experiments in which ipsilateral and contralateral sides were dissociated (Fig. 5), mice with missed injections, bilateral injections (either in ARC or PVH) or low Fos expression in AGRP neurons were excluded.

For bilateral OXT neuron photostimulation following food deprivation, (Supplementary Fig. 14l) experiments were performed in mice bilaterally infected with rAAV2/1-*Oxytocin-ChR2:tdTomato* that were food deprived for 24 h. Before re-feeding (5 min), photostimulation was initiated through a cannula placed over the PVH midline. Mice were allowed *ad libitum* access to food for 2 days after which OXT neurons were photostimulated again for 1 h. Immediately following photostimulation, mice were perfused and their brains were fixed, sectioned, and stained for Oxt- and Fos-immunoreactivity (Supplementary Fig. 14i).

For experiments in which ARC^{AGRP}→PVH and ARC^{AGRP}→PBN projection stimulation were in the same animal, fibres were capped with 1.25 mm OD zirconia ferrules (see <http://syntheticneurobiology.org/protocols/protocoldetail/35/9>), implanted into the brain, and affixed to the skull of the animal with dental

cement. For light delivery, the implanted ferrule-capped fibre was coupled to another optical fibre with a matching 1.25 mm OD zirconium ferrule using a zirconium sleeve.

Pharmacology of AGRP-neuron-evoked feeding. Surgeries and photostimulation were similar to ARC^{AGRP}→PVH stimulation experiments as described above except that cannula placement over the PVH was 0.25 mm lateral to the midline with other coordinates being the same. Prior to photostimulation (−10 min), vehicle (0.15 M saline, DMSO (10%), glacial acetic acid (2.5%)), NPY1R antagonist (BIBO-3304, 3 µg, Tocris), or GABA_A receptor antagonist (bicuculline methiodide, 2.5 pmol, Sigma) were delivered on separate days through the same cannula that was used for fibre implantation. Antagonist injections were on the second or third day, counterbalanced between groups. Injections (50 nl) were through an injection cannula (33 gauge) coupled to a Hamilton syringe that was driven by a Narishige micromanipulator (~30 nl per minute). Animals were subsequently photostimulated (1 h).

PVH pharmacology experiments with the hM3D neuronal activator were performed as above, except that, instead of light delivery, mice were injected with CNO (intraperitoneal, 0.3 mg kg^{−1}, also see below) immediately before intracranial antagonist injection. Two animals had cannula blockage on the final day of injection (condition: bicuculline methiodide followed by photostimulation) and were eliminated from this analysis group.

In separate mice, the effective concentration of BIBO-3304 was determined by the ability to block food intake evoked by NPY (70 pmol, Sigma) injection into the PVH. Consistent with earlier reports⁵¹, NPY-dependent food intake was suppressed. We further confirmed that this concentration does not induce a nonspecific inhibition of overall feeding response, by delivering the same dose into the PVH of 24 h food-deprived animals. Whereas 3 µg BIBO-3304 effectively blocked NPY-induced food intake, it did not significantly decrease re-feeding after deprivation, although there was a trend for reduced food intake (Supplementary Fig. 16).

Neuron silencing. For hM4D-dependent silencing experiments, rAAV2/1-CAG-FLEX-*rev*-hM4D:2a:GFP virus injections were made bilaterally. Mice with total misses or unilateral injections were excluded from analysis after *post hoc* examination of GFP expression. CNO (5 or 0.3 mg kg^{−1}) or saline was delivered by intraperitoneal injection. Control saline injections contained an equivalent amount of DMSO (0.6%). Consistent with a previous report⁵, we found that CNO injection alone did not stimulate food intake in uninfected mice (data not shown).

For PSAM^{L141F}-GlyR silencing experiments in SIM1 neurons, we bilaterally transduced *Sim1-Cre* mice with rAAV-Synapsin-FLEX-*rev*-PSAM^{L141F}-GlyR:IRES:EGFP as described above for hM4D transduction of the PVH. A cognate ligand for this chimaeric chloride channel, PSEM³⁰⁸ (5 mg kg^{−1}) was dissolved in saline and administered intraperitoneally. Food intake was measured before (Pre) and after PSEM³⁰⁸ administration (1 h each). PSEM³⁰⁸ (5 mg kg^{−1}) was also administered to untransduced control mice.

Progressive ratio task. For the entire training protocol, all animals were maintained under *ad libitum* fed conditions, and were never pre-exposed to food deprivation or neuronal manipulation before the days on which these were tested. *AgRP-Cre* or *Sim1-Cre* mice were first acclimated to handling and were exposed to the testing arena in three sessions, where they became familiar with the pellet delivery system and food retrieval. To continue training, each animal had to consume at least 5 pellets from the food hopper on the last session. Animals were then trained to perform lever pressing for food pellets.

The response lever was placed adjacent to the food delivery cup. Animals were allowed one hour in the arena to perform fixed ratio (FR)1, FR3, and FR5, each for 3 days. After five training sessions, animals that were not pressing a lever sufficiently to earn at least three food pellets, were eliminated from further training and testing. To test whether lever pressing was reinforcer-directed, a second identical but inactive lever was placed in the arena on the opposite side of the food delivery cup on the third FR3 session. In successive training and experimental sessions, the inactive lever was rarely pressed as it never led to a food reinforcer.

For AGRP neuron activation experiments, break point testing was performed on a progressive ratio-2 schedule, such that each successive food reward increased the lever-press schedule by two additional responses. Break point is defined here as the lever-press schedule reached at the end of the one hour session⁵². The following day, an optical fibre was inserted into the guide cannula. The animal was allowed 1 h in its home cage to habituate after handling. After transfer to the behavioural test cage, the progressive ratio schedule was repeated in the *ad libitum* fed animal with photostimulation. The following day, food was removed from the home cage (24 h) with water freely available, followed by a break point test under food deprived conditions. After allowing at least 3 days for *ad libitum* repletion, the break point was tested again.

Progressive ratio experiments for SIM1 neuron silencing were performed similarly, except that each test session was extended to 2 h due to temporal variability for pharmacogenetic experiments following intraperitoneal injection. Due to the increased session length, a PR3 schedule was used to prevent satiation. Mice received saline injections in each test session except the day in which behavioural effect of SIM1 neuron silencing was tested. On that day they received a single dose of CNO (5 mg kg^{−1}, intraperitoneally) immediately before being placed in the test cage. The experimenter was blind to the ChR2 or hM4D expression of the subjects, which was revealed after *post hoc* histology.

Image analysis of PVH-selectivity for SIM1 neuron transduction with hM4D was in Image-J. Confocal images across the rostro-caudal axis, included the PVH as well as neighbouring structures (1.3 mm × 1.3 mm), were subjected to the automated thresholding function, and total thresholded surface area was measured. Fluorescence within and outside of the PVH was calculated ($n = 6$ mice). hM4D-transduced SIM1 neurons were primarily located within the PVH ($84 \pm 2\%$ of total fluorescence, $n = 6$ mice), and there was a negative correlation between break point and the small proportion of scattered hM4D-expressing neurons that extended outside the PVH ($r = -0.7$, Supplementary Fig. 13g).

Antibodies. Anti-AGRP (1:5,000, goat, NeuroNomics), anti-AGRP (1:2,000, rat, NeuroNomics), anti-Fos (1:5,000, rabbit, Santa Cruz), anti- α MSH (1:500, sheep, Millipore), anti-POMC (1:200, rabbit, Phoenix Pharmaceuticals), anti-oxytocin (1:3,000, mouse, Abcam), anti-synapsin I (1:1,000, rabbit, Millipore), anti-TdTomato (1:20,000, guinea pig, Covance), anti vGat (1:2,000, rabbit, SySy). Fluorophore-conjugated secondary antibodies were from Invitrogen and Jackson Immuno. Antibodies were diluted in phosphate buffered saline, 1% BSA, 0.1% Triton X-100.

Immunohistochemistry and imaging. After mice were used for behavioural experiments, they were transcardially perfused with 4% paraformaldehyde 0.1 M phosphate buffer fixative. Tissue was post-fixed in this solution for 4–5 h and washed overnight in phosphate buffered saline (pH 7.4). Brain sections (50 µm) were processed for immunohistochemistry, mounted on glass slides using VECTASHIELD mounting medium with 4',6-diamidino-2-phenylindole (DAPI), and coverslipped for imaging.

Photostimulation experiments were confirmed by post hoc Fos quantification. After experiments were complete, mice were photostimulated (1 h) in the absence of food and immediately processed for perfusion, followed by sectioning as described above. Transduction of AGRP, SIM1 or OXT neurons was evaluated by the expression of ChR2-tdtomato. After anti-Fos immunohistochemistry, nuclei were stained with DAPI, and confocal images (2 µm thickness, 5 images) were collected using a 20× (0.8 N.A.) objective, and ChR2 neurons were counted in a single section at the centre of the stack (stacks above and below were also examined to minimise false-negative reporting of Fos expression). Only cells that clearly had a nucleus demarcated by the presence of DAPI staining and surrounded by membrane-localized ChR2-tdtomato fluorescence were included. A subset of those neurons also had Fos-immunoreactivity overlapping with DAPI, and these neurons were taken as Fos-positive ChR2 neurons. Large DAPI-positive nuclei without any surrounding tdtomato fluorescence were considered ChR2-negative neurons. For each mouse, Fos-positive counts from three sections were averaged across the rostral-caudal axis of the ARC or PVH.

Axonal ChR2-penetrance. To quantify axonal ChR2-penetrance for PVH/PBN stimulation experiments, brain slices were immunostained for TdTomato to enhance detection, and confocal images were collected. The percentage of AGRP-containing varicosities transduced with ChR2:tdtomato in the PVH and the PBN were calculated using automated varicosity-detection in Vaa3D⁵³ (confirmed by manual inspection). At least 300 varicosities from three distinct sections along the rostral-caudal axis of the PVH and the PBN were analysed from each animal.

Statistics. Values are represented as means \pm s.e.m. P values for pair-wise comparisons were calculated by two-tailed Student's t -test. P values for comparisons across more than two groups were adjusted with the Holm-Sidak correction. Linear regressions and tests involving one-way and two-way ANOVA with one factor repetition were calculated with SigmaPlot (Systat). n.s. $P > 0.05$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

45. Kaelin, C. B., Xu, A. W., Lu, X. Y. & Barsh, G. S. Transcriptional regulation of *agouti-related protein* (*AgRP*) in transgenic mice. *Endocrinology* **145**, 5798–5806 (2004).
46. Balthasar, N. *et al.* Leptin receptor signaling in POMC neurons is required for normal body weight homeostasis. *Neuron* **42**, 983–991 (2004).
47. Young, W. S. III *et al.* Transgenic expression of green fluorescent protein in mouse oxytocin neurons. *J. Neuroendocrinol.* **11**, 935–939 (1999).
48. Pinto, S. *et al.* Rapid rewiring of arcuate nucleus feeding circuits by leptin. *Science* **304**, 110–115 (2004).
49. Erickson, J. C., Clegg, K. E. & Palmiter, R. D. Sensitivity to leptin and susceptibility to seizures of mice lacking neuropeptide Y. *Nature* **381**, 415–418 (1996).

50. Aravanis, A. M. *et al.* An optical neural interface: *in vivo* control of rodent motor cortex with integrated fiberoptic and optogenetic technology. *J. Neural Eng.* **4**, S143–S156 (2007).
51. Wieland, H. A., Engel, W., Eberlein, W., Rudolf, K. & Doods, H. N. Subtype selectivity of the novel nonpeptide neuropeptide Y Y1 receptor antagonist BIBO 3304 and its effect on feeding in rodents. *Br. J. Pharmacol.* **125**, 549–555 (1998).
52. Zhang, M., Balmadrid, C. & Kelley, A. E. Nucleus accumbens opioid, GABAergic, and dopaminergic modulation of palatable food motivation: contrasting effects revealed by a progressive ratio study in the rat. *Behav. Neurosci.* **117**, 202–211 (2003).
53. Peng, H., Ruan, Z., Atasoy, D. & Sternson, S. Automatic reconstruction of 3D neuron structures using a graph-augmented deformable model. *Bioinformatics* **26**, i38–i46 (2010).

Gut microbiota composition correlates with diet and health in the elderly

Marcus J. Claesson^{1,2*}, Ian B. Jeffery^{1,2*}, Susana Conde³, Susan E. Power¹, Eibhlís M. O'Connor^{1,2}, Siobhán Cusack¹, Hugh M. B. Harris¹, Mairead Coakley⁴, Bhuvaneswari Lakshminarayanan⁴, Orla O'Sullivan⁴, Gerald F. Fitzgerald^{1,2}, Jennifer Deane¹, Michael O'Connor^{5,6}, Norma Harnedy^{5,6}, Kieran O'Connor^{6,7,8}, Denis O'Mahony^{5,6,8}, Douwe van Sinderen^{1,2}, Martina Wallace⁹, Lorraine Brennan⁹, Catherine Stanton^{2,4}, Julian R. Marchesi¹⁰, Anthony P. Fitzgerald^{3,11}, Fergus Shanahan^{2,12}, Colin Hill^{1,2}, R. Paul Ross^{2,4} & Paul W. O'Toole^{1,2}

Alterations in intestinal microbiota composition are associated with several chronic conditions, including obesity and inflammatory diseases. The microbiota of older people displays greater inter-individual variation than that of younger adults. Here we show that the faecal microbiota composition from 178 elderly subjects formed groups, correlating with residence location in the community, day-hospital, rehabilitation or in long-term residential care. However, clustering of subjects by diet separated them by the same residence location and microbiota groupings. The separation of microbiota composition significantly correlated with measures of frailty, co-morbidity, nutritional status, markers of inflammation and with metabolites in faecal water. The individual microbiota of people in long-stay care was significantly less diverse than that of community dwellers. Loss of community-associated microbiota correlated with increased frailty. Collectively, the data support a relationship between diet, microbiota and health status, and indicate a role for diet-driven microbiota alterations in varying rates of health decline upon ageing.

The gut microbiota is required for development and for homeostasis in adult life. Compositional changes have been linked with inflammatory and metabolic disorders¹, including inflammatory bowel disease^{2,3}, irritable bowel syndrome^{4,5} and obesity⁶ in adults. The composition of the human intestinal microbiota is individual-specific at the level of operational taxonomic units (OTUs) and stable over time in healthy adults⁷. The composition of the intestinal microbiota in older people (>65 years) is extremely variable between individuals⁸, and differs from the core microbiota and diversity levels of younger adults^{8,9}. A feature of the ageing process is immunosenescence, evidenced by persistent NF- κ B-mediated inflammation and loss of naive CD4⁺ T cells¹⁰. The microbiota is pivotal for homeostasis in the intestine¹¹, and chronic activation of the innate and adaptive immune system is linked to immunosenescence¹². Correlations have previously been made between specific components of the microbiota and pro-inflammatory cytokine levels, but these did not separate young adults from older people⁹. Alterations in the microbiota composition have also been associated with frailty¹³, albeit in a small cohort from a single residence location.

Deterioration in dentition, salivary function, digestion and intestinal transit time¹⁴ may affect the intestinal microbiota upon ageing. A controllable environmental factor is diet, which has been shown to influence microbiota composition in animal models, in small-scale human studies^{15–20} and over the longer term²¹. However, links between diet, microbiota composition and health in large human cohorts are unclear. To test the hypothesis that variation in the intestinal microbiota of older subjects has an impact on immunosenescence and frailty across the community, we determined the faecal microbiota composition in 178 older people. We also collected dietary intake information, and measured a range of physiological, psychological

and immunological parameters. Dietary groupings were associated with separations in the microbiota and health data sets; the healthiest people live in a community setting, eat differently and have a distinct microbiota from those in long-term residential care. Measures of increased inflammation and increased frailty support a diet-microbiota link to these indicators of accelerated ageing, and suggest how dietary adjustments could promote healthier ageing by modulating the gut microbiota.

Microbiota and residence location

We previously identified considerable inter-individual variability in the faecal microbiota composition of 161 older people (≥ 65 years), including 43 receiving antibiotics⁸. To investigate links between diet, environment, health and microbiota, we analysed 178 subjects, non-antibiotic-treated, for whom we also had dietary information, and stratified by community residence setting: (1) community-dwelling, $n = 83$; (2) attending an out-patient day hospital, $n = 20$; (3) in short-term (< 6 weeks) rehabilitation hospital care, $n = 15$; (4) in long-term residential care (long-stay), $n = 60$. The mean subject age was 78 (± 8 s.d.) years, with a range of 64 to 102 years, and all were of Caucasian (Irish) ethnicity. We included 13 young adults with a mean age of 36 (± 6 s.d.) years. We generated 5.4 million sequence reads from 16S rRNA gene V4 amplicons, with an average of 28,099 ($\pm 10,891$ s.d.) reads per subject.

UniFrac β -diversity analysis indicates the extent of similarity between microbial communities²². UniFrac PCoA (principal co-ordinate) analysis of 47,563 OTUs (grouped at 97% sequence identity) indicated a clear separation between community-dwelling and long-stay subjects using both weighted and un-weighted analysis (Fig. 1a, b). Microbiota from the 13 younger controls clustered with

¹Department of Microbiology, University College Cork, Ireland. ²Alimentary Pharmabiotic Centre, University College Cork, Ireland. ³Department of Statistics, University College Cork, Ireland. ⁴Teagasc, Moorepark Food Research Centre, Moorepark, Fermoy, Co. Cork, Ireland. ⁵Cork University Hospital, Wilton, Cork, Ireland. ⁶St. Finbarr's Hospital, Douglas Road, Cork, Ireland. ⁷Mercy University Hospital, Grenville Place, Cork, Ireland. ⁸South Infirmary, Victoria University Hospital, Cork, Ireland. ⁹Institute of Food and Health, University College Dublin, Ireland. ¹⁰School of Biosciences, Cardiff University, Museum Avenue, Cardiff CF10 3AT, UK. ¹¹Department of Epidemiology and Public Health, University College Cork, Ireland. ¹²Department of Medicine, University College Cork, Ireland.

*These authors contributed equally to this work.

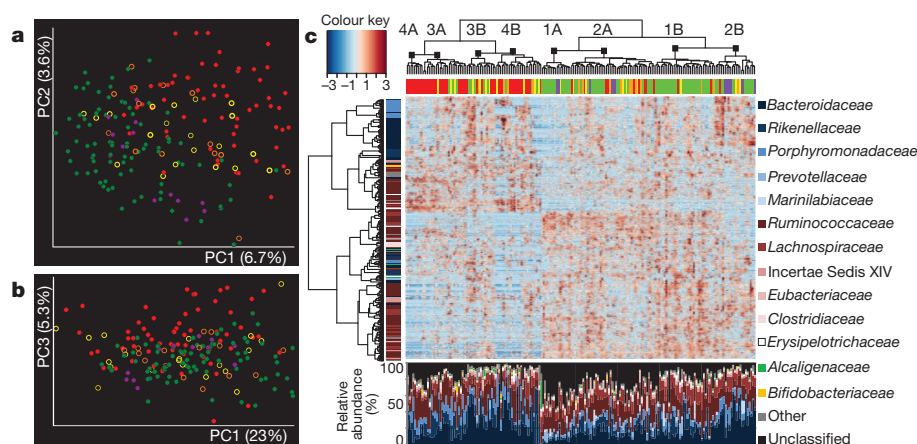


Figure 1 | Microbiota analysis separates elderly subjects based upon where they live in the community. **a**, Unweighted and **b**, weighted UniFrac PCoA of faecal microbiota from 191 subjects. Subject colour coding: green, community; yellow, day hospital; orange, rehabilitation; red, long-stay; and purple, young healthy control subjects. **c**, Hierarchical Ward-linkage clustering based on the Spearman correlation coefficients of the proportion of OTUs, filtered for OTU subject prevalence of at least 20%. Subjects colour coding as in **a**. Labelled

community-dwelling subjects. Eighteen other non-UniFrac β -diversity metrics supported microbiota separation by residence location (Supplementary Fig. 1).

When we examined OTU abundance, we identified a cluster comprised of the majority of the long-stay subjects, separated from the majority of the community-dwelling and young healthy subjects (Fig. 1c). Family-level microbiota assignments showed that long-stay microbiota had a higher proportion of phylum *Bacteroidetes*, compared to a higher proportion of phylum *Firmicutes* and unclassified reads in community-dwelling subjects (Fig. 1c). At genus level, *Coprococcus* and *Roseburia* (of the *Lachnospiraceae* family) were more abundant in the faecal microbiota of community-dwelling subjects (Supplementary Table 1 shows complete list of genera differentially abundant by community location). Genera associated with long-stay subjects included

Parabacteroides, *Eubacterium*, *Anaerotruncus*, *Lactonifactor* and *Coprobacillus* (Supplementary Table 2). The genera associated with community belonged to fewer families, *Lachnospiraceae* were the most dominant. Thus, the microbiota composition of an individual segregated depending on where they lived within a single ethnogeographic region, in a homogeneous cohort where confounding effects of climate, culture, nationality and extreme environment were not a factor.

Concordance of diet and microbiota

Dietary data (for 168 of the 178 subjects, plus five percutaneous endoscopic gastrostomy (PEG)-fed subjects) was collected through a semiquantitative, 147-item, food frequency questionnaire (FFQ), weighted by 10 consumption frequencies. The data were visualized with correspondence analysis (CoA; Fig. 2a). The first CoA axis

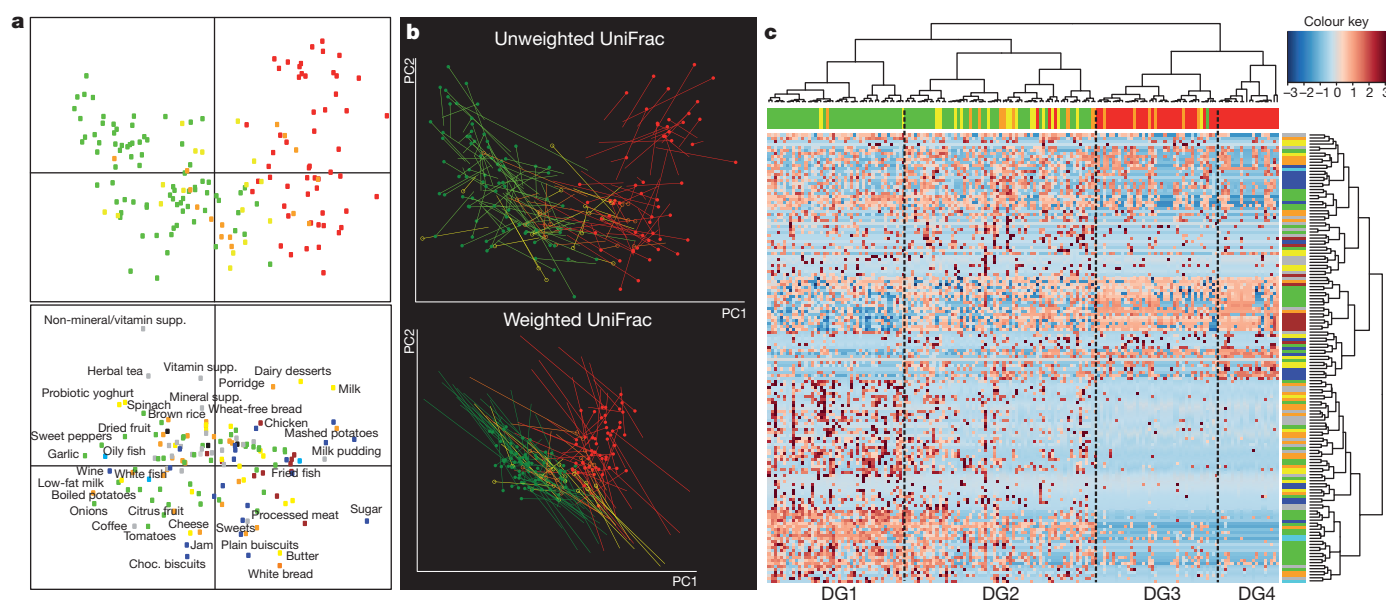


Figure 2 | Dietary patterns in community location correlate with separations based on microbiota composition. **a**, Food correspondence analysis. Top panel, FFQ PCA; bottom panel, driving food types. **b**, Procrustes analysis combining unweighted and weighted UniFrac PCoA of microbiota (non-circle end of lines) with food type PCA (circle-end of lines). **c**, Four dietary groups (DG1, DG2, DG3 and DG4) revealed through complete linkage clustering Euclidean distances applied to first eigenvector in

correspondence analysis. Colour codes in **a**, and horizontal clustering in **b** and **c**, are community location, as per Fig. 1. Food labelling in lower panel in **a**, and vertical clustering in **c**: green, fruit and vegetables; orange, grains such as potatoes, cereals and bread; brown, meat; cyan, fish; yellow, dairy products; blue, sweets, cakes and alcohol; grey vitamins, minerals and tea. Only peripheral and most driving foods are labelled; for a complete list see Supplementary Table 2.

described over 11% of the data set variance and most differences in food consumption between community-dwelling and long-stay subjects. The most discriminating food types were vegetables, fruit and meat, whose consumption changed in a gradual manner along the first eigenvector. Procrustes analysis of the FFQ and the microbiota β -diversity was used to co-visualize the data (Fig. 2b). Separations based on either diet or microbiota co-segregated along the first axis of both data sets (unweighted and weighted UniFrac, Fig. 2b; Monte-Carlo P value < 0.0001). Application of complete linkage clustering and Euclidean distances to the first eigenvector (Fig. 2c) revealed four dietary groups (DGs). DG1 ('low fat/high fibre') and DG2 ('moderate fat/high fibre') included 98% of the community and day hospital subjects, and DG3 ('moderate fat/low fibre') and DG4 ('high fat/low fibre') included 83% of the long-stay subjects. For a complete description of dietary groups, see Supplementary Notes and Supplementary Table 3.

The healthy food diversity index (HFD²³) positively correlated with three microbiota diversity indices (Supplementary Fig. 2a), and all four indices showed significant differences between community and long-stay subjects (Supplementary Fig. 2b), indicating that a healthy, diverse diet promotes a more diverse gut microbiota. Analysing by dietary groups rather than residence location confirmed that both microbiota and diet were most diverse in DG1, and least diverse in DG3 and DG4 (Supplementary Fig. 3). Procrustes analysis similarly showed that the dietary groups were associated with separations in microbiota composition (Supplementary Fig. 3). Furthermore, the microbiota was associated with the duration in long-stay, with residents of more than a year having a microbiota that was furthest separated from community-dwelling subjects (Supplementary Fig. 4). For the majority of these longer-term residents, the diet was different from that in more recently admitted subjects (Supplementary Fig. 4). Examination of duration of care (Supplementary Fig. 4c) showed that diet changed more quickly than the microbiota did; both diet and microbiota moved in the direction away from the community types. After 1 month in long stay, all subjects had a long-stay diet, but it took a year for the microbiota to be clearly the long-stay type. Collectively the data indicate that the composition of the microbiota is determined by the composition and diversity of the diet.

Community setting and faecal metabolome

Faecal metabolites correlate with microbiota composition and inflammatory scores in Crohn's disease²⁴. We therefore performed metabolomic analysis (NMR spectroscopy) of faecal water from 29

subjects, representative (by UniFrac) of three community settings. (Day-hospital subjects grouped closely to community dwellers by microbiota and dietary analysis, and were not included.) A representative NMR profile is presented in Supplementary Fig. 5. Initial PCA (principal component analysis) analysis showed a trend for separation according to community setting (data not shown). Pair-wise statistical models were therefore constructed according to the cluster groups. Valid and robust models were obtained for comparison of NMR spectra from community and long-stay subjects, and community and rehabilitation subjects (Fig. 3). The major metabolites separating community from long-stay subjects were glucose, glycine and lipids (higher levels in long-stay than community subjects), and glutarate and butyrate (higher levels in community subjects). Co-inertia analysis of the genus-level microbiota and metabolome data revealed a significant relationship (P value < 0.01) between the two data sets (Supplementary Fig. 6 and Supplementary Notes). Notwithstanding three long-stay subjects, a diagonal separated community from long-stay in both microbiota and metabolome data sets. Other metabolites of interest were acetate, propionate and valerate, which were more abundant in community dwellers (Supplementary Fig. 6).

To investigate microbial short-chain fatty acid (SCFA) production further, the frequency of microbial genes for SCFA production was investigated by shotgun metagenomic sequencing. We sequenced 125.9 gigabases (Gb) of bacterial DNA from 27 of the 29 subjects, and assembled contigs with a total length of 2.20 Gb, containing 2.51 million predicted genes (Supplementary Table 4). Consistent with reduced microbiota diversity (Supplementary Fig. 3), there were significantly fewer total genes predicted, and higher N50 values (N50 is the length of the smallest contig that contains the fewest (largest) contigs whose combined length represents at least 50% of the assembly), in the assembled metagenomic data of long-stay subjects compared to rehabilitation or community subjects (Supplementary Fig. 7). The metagenomes were then searched for key microbial genes in butyrate, acetate and propionate production, revealing significantly higher gene counts and coverage for butyrate- and acetate-producing enzymes (BCoAt and ACS, respectively) in community and rehabilitation compared to long-stay subjects (Supplementary Fig. 8 and Supplementary Table 5). There was also significantly higher coverage of the propionate-related genes (PCoAt) in community compared to long-stay subjects, but the higher gene count was not significant (Supplementary Table 5). These observations are consistent with the association of butyrate, acetate and propionate and the direction of the main split between long-stay and community subjects in the metabolome; candidate

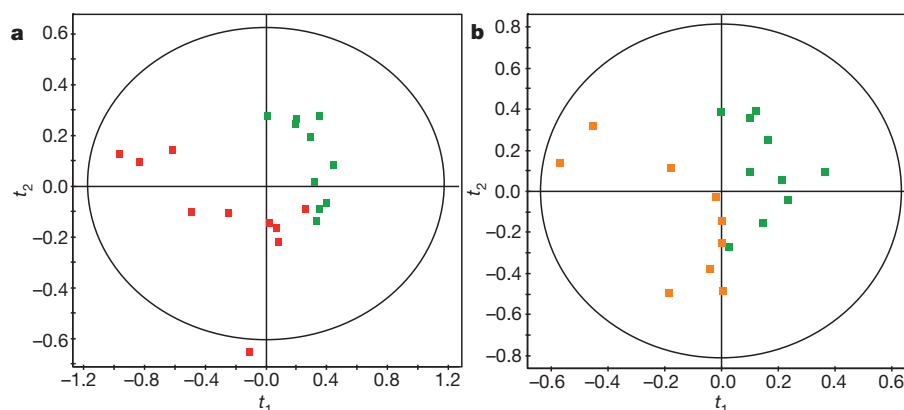


Figure 3 | PLS-DA plots of ^1H NMR spectra of faecal water from community, long-stay and rehabilitation subjects. **a**, Community subjects (green) versus long-stay subjects (red); $R^2 = 0.517$, $Q^2 = 0.409$, two-component model. **b**, Community subjects (green) versus rehabilitation subjects (orange); $R^2 = 0.427$, $Q^2 = 0.163$, two-component model. The ellipses represent the Hotelling's T^2 with 95% confidence. To confirm the validation of the model, permutation tests ($n = 1,000$) were performed. For model **a**, the 95%

confidence interval for the misclassification error rate (MER) was (0.43, 0.57). Using the PLS-DA model on the data resulted in an MER of 0.2 which is outside the 95% confidence interval obtained for random permutation tests, thus validating the model. For model **b**, using permutation testing the 95% confidence interval for the MER was (0.45, 0.55). Using the PLS-DA model on the data resulted in an MER of 0.16 which is outside the 95% confidence interval obtained for random permutation tests.

genera associated include *Ruminococcus* and *Butyricoccus* for butyrate production (Supplementary Fig. 6), but require validation in larger cohorts. Microbiota function deduced from the metagenome thus corresponded to the measured metabolome for at least one key metabolite that can affect health²⁵.

Microbiota–health correlations

Markers of inflammation (serum TNF- α , IL-6 and IL-8 and C-reactive protein (CRP)) had significantly higher levels in long-stay and rehabilitation subjects than in community dwellers (Supplementary Fig. 9). Long-stay subjects also scored poorly for diverse health parameters (Supplementary Tables 6 and 7), including the Charlson comorbidity index (CCI, a robust predictor of survival encompassing 19 medical conditions²⁶), the geriatric depression test (GDT), the Barthel index²⁷, functional independence measure (FIM²⁸), mini-mental state exam (MMSE²⁹) and mini nutritional assessment (MNA³⁰).

Correlations between health parameters and microbiota composition were examined using quantile (median) regression tests, adjusted for gender, age and community setting with an additive model (Supplementary Methods). Median regression gives less weight to extreme values than the linear regression based on ordinary least squares and consequently, is less influenced by outliers. The model was adjusted for medications that might influence the tested parameters (Supplementary Table 8). The effect of medication was generally small (Supplementary Table 8). Because ethnicity was exclusively Irish Caucasian it did not require model adjustment. The microbiota composition did not differ for males and females after adjusting for age and location.

Significant associations between several health/frailty measurements and the major separations from microbiota UniFrac analysis (Fig. 1) are shown in Table 1. For example, a positive change in

microbiota along the full range of the PC1 axis in the un-weighted UniFrac PCoA for long-stay-only subjects was associated with inflammation (CRP increase of 13.9 mg l⁻¹), and other inflammatory markers significantly correlated with microbiota (IL-6 and IL-8, whole cohort). As expected, there was minimal variability amongst community-dwelling subjects, but within the long-stay subjects the most significant associations were related to functional independence (FIM), Barthel index and nutrition (MNA), followed by blood pressure and calf circumference. The latter may be attributable to the influence of diet and/or the microbiota on muscle mass, sarcopaenia³¹ and thereby on frailty. This was supported by investigation of linkage between frailty and faecal metabolites (probabilistic principal components and covariates analysis; PPCA³²). Thus, the FIM and Barthel indices were significant covariates with the faecal water metabolome (Supplementary Fig. 10) and levels of acetate, butyrate and propionate increased with higher values of both indices (that is, less frail subjects). Among community-dwelling subjects, there was also a strong association between microbial composition and nutrition (MNA) and a weaker link with blood pressure, for which a relationship with the microbiota has previously been established³³. There was no correlation between the *Bacteroidetes:Firmicutes* ratio and body mass index (BMI), although there was a correlation with overall microbiota in long-stay subjects. Measures for the geriatric depression test (GDT) showed significant microbiota association with PCoA axis 2 (Table 1). We detected no significant confounding of microbiota–health correlations due to medications, antibiotic treatment (before the 1-month exclusion window), and diet–health correlations separate from dietary impact on microbiota (Supplementary Notes).

Taken together, the major trends in the microbiota that separated healthy community subjects from less healthy long-stay subjects were

Table 1 | Regression tests of associations between clinical measurements and microbiota composition.

a Unweighted UniFrac PCoA for all four residence locations									
Parameter	PC1			PC2			PC3		
	RC range	RC s.d.	P	RC range	RC s.d.	P	RC range	RC s.d.	P
GDT	–0.42	–0.11	0.6	–2.7	–0.54	0.037	0.18	0.04	0.84
Diastolic blood pressure	0.97	0.25	0.81	–10.1	–2.02	0.033	–14.2	–3.1	0.001
Weight	–14.6	–3.8	0.033	–7.16	–1.43	0.27	–1.57	–7.2	0.18
CC	–3.9	–1.01	0.022	–2.9	–0.58	0.19	–3.2	–0.7	0.047
IL-6	6.71	1.7	0.006	6.1	1.22	0.007	2.08	0.45	0.2
IL-8	4.23	1.1	0.43	13.6	2.7	0.03	4.06	0.89	0.47376716
TNF- α	1.1	0.28	0.31	0.62	0.13	0.72	3.9	0.9	0.0005
b Unweighted UniFrac PCoA for community-only subjects									
Parameter	PC1			PC2			PC3		
	RC range	RC s.d.	P	RC range	RC s.d.	P	RC range	RC s.d.	P
MNA	–1.1	–0.26	0.29	1.9	0.5	0.006	0.7	0.14	0.59
Diastolic blood pressure	–8.4	–1.98	0.08	14.3	3.4	0.035	–15.72	–3.26	0.13
GDT	–0.13	–0.03	0.8	–1.5	–0.35	0.02	–0.8	–0.16	0.4
c Unweighted UniFrac PCoA for long-stay-only subjects									
Parameter	PC1			PC2			PC3		
	RC range	RC s.d.	P	RC range	RC s.d.	P	RC range	RC s.d.	P
Barthel	–6	–1.5	0.004	–4.8	–1.3	0.036	–0.6	–0.15	0.71
FIM	–30.8	–7.8	0.046	–33.3	–4.7	0.024	–2.42	–0.6	0.86
MMSE	–12.15	–3.08	0.14	–18.4	–4.8	0.009	3.22	0.8	0.63
MNA	–3.87	–0.98	0.23	–11.2	–3	0.004	–0.02	–0.005	0.99
BMI	–1.2	–0.31	0.69	–5	–1.3	0.047	–0.24	–0.06	0.92
CC	0.2	0.05	0.93	–6.8	–1.77	0.0016	0.45	0.11	0.82
Diastolic blood pressure	19.3	4.9	0.015	–12.4	–3.24	0.034	–15.4	–3.81	0.007
Systolic blood pressure	36.5	9.3	0.007	–1.57	–0.41	0.83	–2.05	–0.51	0.87
Weight	–3.2	–0.81	0.69	–12.7	–3.3	0.024	–2.48	–0.61	0.72
IL-8	–2.56	–0.65	0.78	22.31	5.84	0.006	1.14	0.28	0.93
CRP	13.9	3.53	0.02	–3.01	–0.8	0.27	–2.54	–0.63	0.61

Quantile (median) regression tests of associations between clinical measurements and microbiota composition as measured by unweighted UniFrac PCoA across all four residence locations (that is, all subjects (a), community-only subjects (b) and long-stay-only subjects (c)). Column headings are: RC range, regression coefficients scaled to the full variation along each PCoA axis, thus indicating relative magnitude and direction of the health association; RC s.d., regression coefficients scaled to one standard deviation; P, quantile regression P values generated by boot-strap analysis. Significant associations are in bold. An additive model was used to adjust for the effects of age, sex, residence location, relevant medication and the two other principal coordinates. CC, calf circumference; IL, interleukin; MMSE, mini-mental state examination.

associated with markers for increased frailty and poorer health, having adjusted for gender, age and location. Because location largely determines diet (Fig. 2), adjusting for location reduces the effect of diet, and as there was also clear evidence for microbiota–health associations within the long-stay setting, we infer that the causal relationship is in a diet–microbiota–health direction.

Microbiota structure and healthy ageing

Gut microbiota can be assigned to one of three enterotypes³⁴, driven by *Bacteroides*, *Prevotella* and *Ruminococcus* species. A recent study detected only the *Bacteroides* and *Prevotella* enterotypes, which were associated with diets rich in protein and carbohydrate, respectively²¹. Using those methods, we predicted an optimal number of two clusters using five out of six methodologies, albeit with weaker support than previous studies (Supplementary Fig. 11). In line with a previous study²¹, the two clusters associated with *Bacteroides* and *Prevotella*, but not with *Ruminococcus*. Although enterotype assignments from the three approaches were very different (Supplementary Fig. 11), community subjects were more frequently of the *Prevotella* enterotype.

To identify patterns in the microbiota, we established co-abundance associations of genera (Supplementary Fig. 12a), and then clustered correlated genera into six co-abundance groups (CAGs) (Supplementary Fig. 12b). These are not alternatives to enterotypes, which are subject-driven and poorly supported in this elderly cohort, but they describe the microbiota structures found across the subject groups in statistically significant co-abundance groups (Supplementary Notes). The dominant genera in these CAGs were *Bacteroides*, *Prevotella*,

Ruminococcus, *Oscillibacter*, *Alistipes* and the central *Odoribacter* CAG. These CAG relationships are termed Wiggum plots, in which genus abundance can be represented as discs proportional to abundance (Supplementary Fig. 12), to normalized over-abundance (Fig. 4), or to differential over-abundance (Supplementary Fig. 13). In the Wiggum plot corresponding to the whole cohort (Supplementary Fig. 12), the path away from the *Ruminococcus* CAG towards the *Oscillibacter* CAG shows a reduced number of genera that make butyrate, and an increased number able to metabolize fermentation products.

To simplify the microbiota data for health correlation, we used the eight subject divisions identified by OTU clustering (Fig. 1c). These eight divisions were superimposed on a UniFrac PCoA analysis of the data in Fig. 1a, defining 8 subject groups (Fig. 4, Groups 1A through 4B). These are separation points within a microbiota composition spectrum that represent groups of individuals who have significantly different microbiota as defined by the permutation multivariate analysis of variance (MANOVA) test on unweighted UniFrac data. We then constructed individual Wiggum plots for the microbiota in these 8 groups (Fig. 4). The transition from healthy community-dwelling subjects, to frail long-term care residents, is accompanied by distinctive CAG dominance, most significantly in abundances of *Prevotella* and *Ruminococcus* CAGs (community associated CAGs) and *Alistipes* and *Oscillibacter* CAGs (long-stay-associated CAGs).

Our analysis of Fig. 4 suggested two paths from community-associated health to long-stay-associated frailty (plot 1A–4A, and 1B–4B), which were examined with reference to health correlations

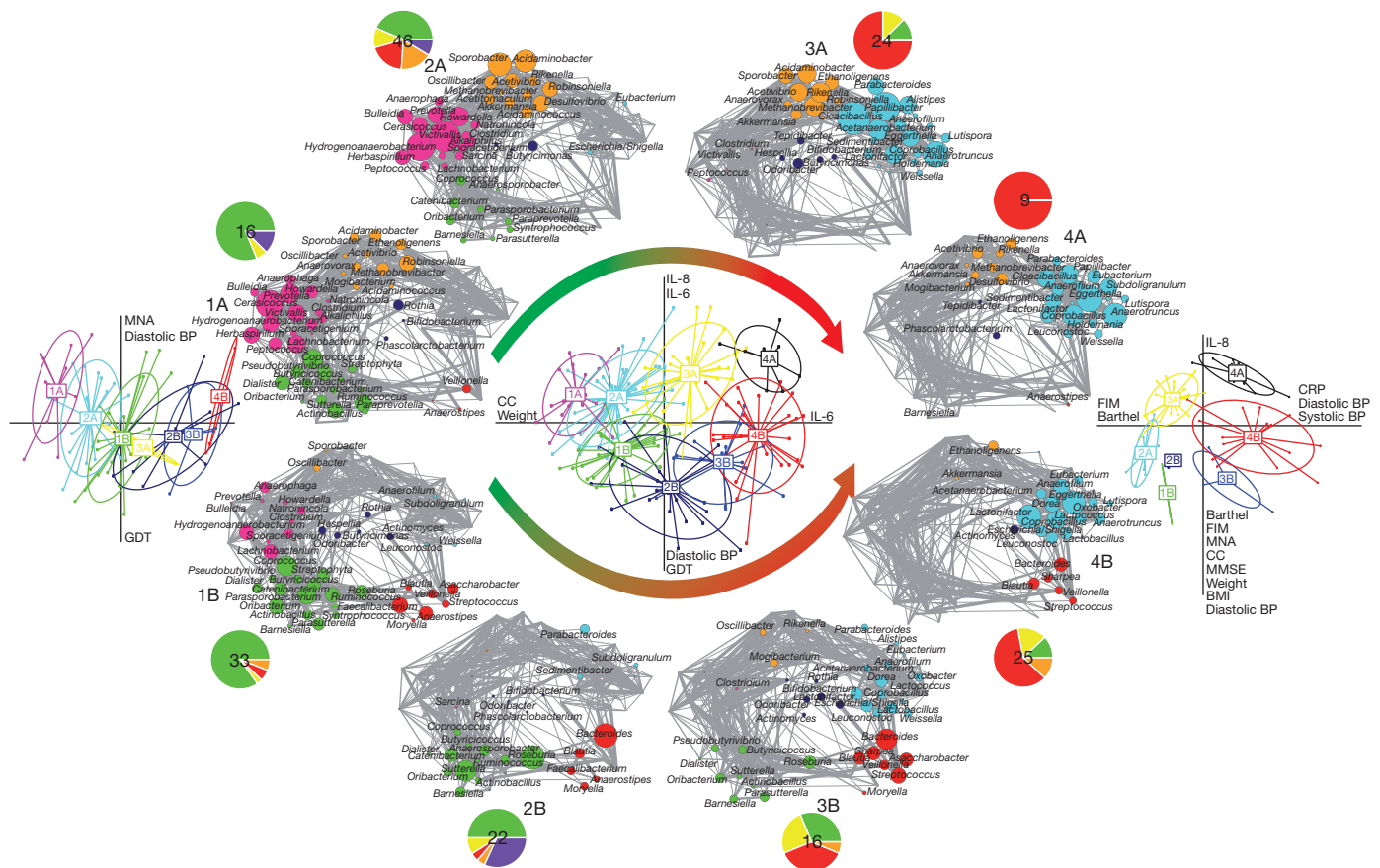


Figure 4 | Transition in microbiota composition across residence location is mirrored by changes in health indices. The PCoA plots show 8 groups of subjects defined by unweighted UniFrac microbiota analysis of community subjects (left), the whole cohort (centre), and long-stay subjects (right). The main circle shows the Wiggum plots corresponding to the 8 groups from whole-cohort analysis, in which disc sizes indicate genus over-abundance

relative to background. The pie charts show residence location proportions (colour coded as in Fig. 1c) and number of subjects per subject group. Curved arrows indicate transition from health (green) to frailty (red). FIM, functional independence measure; MNA, mini nutritional assessment; GDT, geriatric depression test; CC, calf circumference; CRP, C-reactive protein; IL, interleukin; BP, blood pressure; MMSE, mini-mental state examination.

in Table 1, plus separate PCoAs for the community-only, and long-stay-only subjects. The community and whole-cohort analyses identified an association of depression with axis 2—subjects in the lower path had higher GDT scores. IL-6 and IL-8 levels were higher in the upper path by whole-cohort analysis (Fig. 4 and Supplementary Fig. 14), whereas CRP levels were higher in the lower path in long-stay-only analysis. Furthermore, subjects in the lower path had higher systolic and diastolic blood pressure, except in the community-only analysis. This apparent inconsistency is explained by a highly significant change in diastolic blood pressure along the primary PCoA axis in the long-stay subjects, emphasizing the value of a stratified cohort. The subjects in the upper path were older but had higher Barthel and FIM scores than subjects of a similar age in the lower path (Supplementary Fig. 14), consistent with healthier ageing. Movement along PCoA axis 1 of the whole cohort (that is, from community to long-stay, left to right, Fig. 4) is associated with a reduction in abundance of *Ruminococcus* and *Prevotella*, and increased abundance of the *Oscillibacter* CAG, accompanied by calf circumference decrease and weight decrease (Table 1), and increase in IL-6 levels. Moving along axis 1 of the long-stay PCA (that is, between the two right-ward arms), there is a reduction in the *Oscillibacter* CAG, increase in abundance of the *Bacteroides* CAG, reduced FIM and Barthel indices, and increased levels of CRP (Fig. 4). Consideration of the microbiota–health correlations in the long-stay cohort (Fig. 4), upwards along axis 2, highlights the association with increased frailty, reduced muscle mass, and poorer mental activity moving away from community-type microbiota.

Health–microbiota associations were statistically significant, even when regression models were adjusted for location. Although other factors undoubtedly contribute to health decline, and are difficult to completely adjust for in retrospective studies, the most plausible interpretation of our data is that diet shapes the microbiota, which then affects health in older people. Diet-determined differences in microbiota composition may have subtle impacts in young adults in developed countries. These would be difficult to correlate with health parameters, but become far more evident in the elderly who are immunophysiologically compromised. This is supported by the stronger microbiota–health associations evident in the long-stay cohort, and there is now a reasonable case for microbiota-related acceleration of ageing-related health deterioration. An ageing population is now a general feature of western countries^{35,36} and an emerging phenomenon even among developing countries. The association of the intestinal microbiota of older people with inflammation¹² and the clear association between diet and microbiota outlined in this and previous studies^{20,21,37,38} argue in favour of an approach of modulating the microbiota with dietary interventions designed to promote healthier ageing. Dietary supplements with defined food ingredients that promote particular components of the microbiota may prove useful for maintaining health in older people. On a community basis, microbiota profiling, potentially coupled with metabolomics, offers the potential for biomarker-based identification of individuals at risk for, or undergoing, less-healthy ageing.

METHODS SUMMARY

Amplicons of the 16S rRNA gene V4 region were sequenced on a 454 Genome Sequencer FLX Titanium platform. Sequencing reads were quality filtered, OTU clustered, ChimeraSlayer filtered and further analysed using the QIIME pipeline³⁹ and RDP-classifier⁴⁰. Statistical analysis was performed using Stata and R software packages. Nuclear magnetic resonance (NMR) spectroscopy was performed on a 600 MHz Varian NMR Spectrometer as previously described⁴¹.

Habitual dietary intake was assessed using a validated, semiquantitative, FFQ, administered by personnel who received standardized training in dietary assessment. FFQ coding, data cleaning and data checks were conducted by a single, trained individual to ensure consistency of data.

Full Methods and any associated references are available in the online version of the paper.

Received 10 January; accepted 14 June 2012.

Published online 13 July 2012.

- O'Toole, P. W. & Claesson, M. J. Gut microbiota: changes throughout the lifespan from infancy to elderly. *Int. Dairy J.* **20**, 281–291 (2010).
- Frank, D. N. *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl Acad. Sci. USA* **104**, 13780–13785 (2007).
- Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
- Kassinen, A. *et al.* The fecal microbiota of irritable bowel syndrome patients differs significantly from that of healthy subjects. *Gastroenterology* **133**, 24–33 (2007).
- Jeffery, I. B. *et al.* An irritable bowel syndrome subtype defined by species-specific alterations in faecal microbiota. *Gut* **61**, 997–1006 (2012).
- Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022–1023 (2006).
- Rajilić-Stojanović, M. *et al.* Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. *Environ. Microbiol.* **11**, 1736–1751 (2009).
- Claesson, M. J. *et al.* Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc. Natl Acad. Sci. USA* **108** (Suppl 1), 4586–4591 (2011).
- Biagi, E. *et al.* Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. *PLoS ONE* **5**, e10667 (2010).
- Franceschi, C. *et al.* Inflamm-aging: an evolutionary perspective on immunosenescence. *Ann. NY Acad. Sci.* **908**, 244–254 (2000).
- Garrett, W. S., Gordon, J. I. & Glimcher, L. H. Homeostasis and inflammation in the intestine. *Cell* **140**, 859–870 (2010).
- Guigoz, Y., Dore, J. & Schiffrin, E. J. The inflammatory status of old age can be nurtured from the intestinal environment. *Curr. Opin. Clin. Nutr. Metab. Care* **11**, 13–20 (2008).
- van Tongeren, S. P., Slaets, J. P., Harmsen, H. J. & Welling, G. W. Fecal microbiota composition and frailty. *Appl. Environ. Microbiol.* **71**, 6438–6442 (2005).
- Lovat, L. B. Age related changes in gut physiology and nutritional status. *Gut* **38**, 306–309 (1996).
- Hildebrandt, M. A. *et al.* High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology* **137**, 1716–1724 (2009).
- Mai, V., McCrary, Q. M., Sinha, R. & Gleis, M. Associations between dietary habits and body mass index with gut microbiota composition and fecal water genotoxicity: an observational study in African American and Caucasian American volunteers. *Nutr. J.* **8**, 49 (2009).
- Muegge, B. D. *et al.* Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332**, 970–974 (2011).
- De Filippo, C. *et al.* Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl Acad. Sci. USA* **107**, 14691–14696 (2010).
- Turnbaugh, P. J. *et al.* The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci. Transl. Med.* **1**, 6ra14 (2009).
- Faith, J. J., McNulty, N. P., Rey, F. E. & Gordon, J. I. Predicting a human gut microbiota's response to diet in gnotobiotic mice. *Science* **333**, 101–104 (2011).
- Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–108 (2011).
- Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**, 8228–8235 (2005).
- Drescher, L. S., Thiele, S. & Mensink, G. B. A new index to measure healthy food diversity better reflects a healthy diet than traditional measures. *J. Nutr.* **137**, 647–651 (2007).
- Jansson, J. *et al.* Metabolomics reveals metabolic biomarkers of Crohn's disease. *PLoS ONE* **4**, e6386 (2009).
- Pryde, S. E., Duncan, S. H., Hold, G. L., Stewart, C. S. & Flint, H. J. The microbiology of butyrate formation in the human colon. *FEMS Microbiol. Lett.* **217**, 133–139 (2002).
- de Groot, V., Beckerman, H., Lankhorst, G. J. & Bouter, L. M. How to measure comorbidity: a critical review of available methods. *J. Clin. Epidemiol.* **56**, 221–229 (2003).
- Mahoney, F. I. & Barthel, D. W. Functional evaluation: the Barthel index. *Md. State Med. J.* **14**, 61–65 (1965).
- Kidd, D. *et al.* The functional independence measure: a comparative validity and reliability study. *Disabil. Rehabil.* **17**, 10–14 (1995).
- Folstein, M. F., Folstein, S. E. & McHugh, P. R. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **12**, 189–198 (1975).
- Bauer, J. M., Kaiser, M. J., Anthony, P., Guigoz, Y. & Sieber, C. C. The mini nutritional assessment—its history, today's practice, and future perspectives. *Nutr. Clin. Pract.* **23**, 388–396 (2008).
- Cruz-Jentoft, A. J. *et al.* Sarcopenia: European consensus on definition and diagnosis: report of the European Working Group on Sarcopenia in Older People. *Age Ageing* **39**, 412–423 (2010).
- Nyamundanda, G., Brennan, L. & Gormley, I. C. Probabilistic principal component analysis for metabolomic data. *BMC Bioinformatics* **11**, 571 (2010).

33. Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011).
34. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
35. European Commission. *Population structure and ageing* http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Population_structure_and_ageing (2011).
36. Kinsella, K. & He, W. *An Aging World: 2008* (US Government Printing Office, 2009).
37. Kau, A. L., Ahern, P. P., Griffin, N. W., Goodman, A. L. & Gordon, J. I. Human nutrition, the gut microbiome and the immune system. *Nature* **474**, 327–336 (2011).
38. Walker, A. W. *et al.* Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *ISME J.* **5**, 220–230 (2011).
39. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**, 335–336 (2010).
40. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
41. O'Sullivan, A., Gibney, M. J. & Brennan, L. Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. *Am. J. Clin. Nutr.* **93**, 314–321 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Government of Ireland National Development Plan by way of a Department of Agriculture Food and Marine, and Health Research Board FHRI award to the ELDERMET project, as well as by a Science Foundation Ireland award to the Alimentary Pharmabiotic Centre. M.J.C. is funded by a fellowship from the Health Research Board of Ireland. We thank K. O'Donovan and P. Egan for clinical assistance, staff in Cork City and County hospitals for facilitating subject recruitment, S. Wong and B. Clayton for supercomputer access.

Author Contributions All authors are members of the ELDERMET consortium (<http://eldermet.ucc.ie>). P.W.O.T., E.M.O.C., S.Cu.¹ and R.P.R. managed the project; D.v.S., G.F.F., C.S., J.R.M., F.S., C.H., R.P.R. and P.W.O.T. designed the analyses; M.J.C., I.B.J., S.Co.³, E.M.O.C., H.M.B.H., M.C., B.L., O.O.S., A.P.F., S.E.P., M.W. and L.B. performed the analyses; J.D. performed DNA extraction and PCR; M.W. and L.B. performed NMR metabolomics; M.O.C., N.H., K.O.C. and D.O.M. performed clinical analyses; M.J.C., I.B.J., S.Co.³, E.M.O.C., L.B., J.R.M., A.P.F., R.P.R., C.H., F.S. and P.W.O.T. wrote the manuscript.

Author Information Amplicon sequence data, and shotgun sequence data, contigs, genes and annotations, have been deposited in MG-RAST under the Project ID 154 (<http://metagenomics.anl.gov/linkin.cgi?project154>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.W.O.T. (pwotoole@ucc.ie).

METHODS

Subject recruitment and sample collection. This study was approved by the Cork Clinical Research Ethics Committee. Subjects older than 64 years were recruited and clinically investigated in two local hospitals, which serve a population base of ~481,000 in the Cork city and county region. They were defined as (1) community-dwelling (community); (2) attending an out-patient day hospital (out-patient); (3) in short-term rehabilitation hospital care (rehabilitation; under 6 weeks stay) or (4) in long-term institutionalized care (long stay; more than 6 weeks). The mean age of the subjects was 78 (± 8) years, with a range of 64 to 102 years. The subjects were all of Irish ethnicity. None of the faecal samples from elderly subjects from our previous study⁸ were analysed in the current analysis, because we did not have food frequency data for all that cohort. Exclusion criteria were a history of alcohol abuse, participation in an investigational drug evaluation or antibiotic treatment within the previous 30 days, or advanced organic disease. Informed consent was obtained from all subjects or, in cases of cognitive impairment, by next-of-kin in accordance with the local research ethics committee guidelines. Data collected included anthropometric measurements, clinical history and status and medication history. Antibiotic use before the one-month exclusion period was also recorded for each subject. Thirteen younger adult subjects of age ranging 28–46 years, which had not been treated with antibiotics within 30 days, were also recruited by informed consent.

Clinical and nutritional data collection. Habitual dietary intake was assessed using a validated, semi-quantitative, food frequency questionnaire (FFQ) based upon the SLAN study⁴². Food properties were determined using the UK Food Standards Agency Nutrient databank⁴³. The mini nutritional assessment (MNA) was used as a screening and assessment tool to identify subjects at risk of malnutrition.

Non-fasted blood samples were collected and analysed at Cork University Hospital clinical laboratories. Cytokines were measured using validated, commercial multi-spot microplates (Meso Scale Diagnostics). Anthropometric measures included height, weight, calf and mid-arm circumference. Charlson comorbidity index, mini mental state exam, geriatric depression test, Barthel score and functional independence measures were carried out on all participants. For long-term care, day-hospital and rehabilitation subjects, a research nurse reviewed the medical records for information on disease and current medication usage.

Molecular methods and bioinformatics. DNA was extracted from faecal samples, and the V4 region of the 16S rRNA gene was amplified, sequenced and analysed, as described previously⁴⁴. Briefly, V4 amplicons were sequenced on a 454 Genome Sequencer FLX Titanium platform (Roche Diagnostics and Beckman Coulter Genomics). Raw sequencing reads were quality trimmed using the QIME pipeline³⁹ according to the following criteria: (1) exact matches to primer sequences and barcode tags; (2) no ambiguous bases (Ns); (3) read-lengths not shorter than 150 base pairs (bp) or longer than 350 bp; (4) the average quality score in a sliding window of 50 bp not to fall below 25. For large-scale assignments into the new Bergey's bacterial taxonomy⁴⁵ we used the RDP-classifier version 2.2 with 50% as confidence value threshold. This was based on what was found suitable for V4 amplicons from the human gut environment⁴⁴. RDP classifications were imported into a MySQL database for efficient storage and advanced querying.

The amplicon reads were clustered into OTUs at 97% identity level, and filtered for chimeric sequences using ChimeraSlayer (http://microbiomeutil.sourceforge.net/#A_CS). Representative sequences (the most abundant) for each OTU were aligned using PyNAST⁴⁶ before tree building using FastTree⁴⁷. These phylogenies were combined with absence/presence or abundance information for each OTU to calculate unweighted or weighted UniFrac distances, respectively⁴⁸. Principal coordinate analysis and Procrustes superimposition were then performed from the UniFrac distances and Food Frequency data. The amplicon sequences were deposited in MG-RAST under the Project ID 154.

Metagenomes were sequenced from libraries with 91 bp paired-end Illumina reads and 350 bp insert size and assembled using MetaVelvet⁴⁹. Samples EM039 and EM173 were sequenced from libraries of 101 bp paired-end Illumina reads with a 500 bp insert size, and subsequently assembled using MIRA⁵⁰ in hybrid with 551,726 and 665,164 454 Titanium reads, respectively. Protein sequences from enzymes were screened against the assembled metagenomes using TBLASTN with an amino acid identity cut-off of 30% and an alignment length cut-off of 200 bp. We screened the metagenome data for enzymes associated with production of butyrate (butyryl-CoA transferase/acetyl-CoA hydrolase), acetate (acetate-formyltetrahydrofolate synthetase/formate-tetrahydrofolate ligase), and propionate (propionyl-CoA:succinate-CoA transferase/propionate CoA-transferase). Genes were predicted using MetaGene⁵¹.

NMR analysis of the faecal water metabolome. Faecal water samples were prepared by the addition of 60 μ l D₂O and 10 μ l tri-methylsilyl-2,2,3,3-tetradeuterio-propionate to 540 μ l faecal water. Spectra of samples were acquired by using a Carr–Purcell–Meiboom–Gill (CPMG) pulse sequence with 32k data points and 256 scans. Spectra were referenced to TSP at 0.0 p.p.m., phase and baseline

corrected with a line broadening of 0.3 Hz using the processor on Chenomx NMR suite 7 (Chenomx). The spectra were integrated at full resolution for data analysis (PCA, PLS-DA, CIA) with the water region (4–6 p.p.m.) excluded and the data was normalized to the sum of the spectral integral. For PPCCA data analysis, the spectra were integrated into spectral regions (0.01 p.p.m.). Two-dimensional ¹H–¹H correlation spectroscopy (COSY) and total correlation spectroscopy (TOCSY) were acquired on a 600 MHz NMR spectrometer. TOCSY spectra were acquired with a spin lock of 65 ms. All two-dimensional data were recorded with standard Varian pulse sequences collecting 1,024 \times 128 data points with a sweep width of 9.6 kHz and 32 scans per increment.

Statistical methods and metabolome data analysis. Statistical analysis was carried out using R (version 2.13.2) or Stata (version 11) software packages. Kruskal–Wallis and Mann–Whitney tests were used to find significant differences in microbial taxa, clinical and biochemical measures, alpha diversity, and Healthy Food Diversity (HFD). Data were visualized by boxplots. Unless stated otherwise, box plots represented the median and interquartile ranges, with the error bars showing the last datum within 1.5 of the interquartile range of the upper and lower quartiles. We used least square linear regression for comparing alpha diversity and HFD. Median regression⁵² was used to compare clinical measures and microbiota, while adjusting for age, gender, medications, and when appropriate residence location. For median regression, the median was modelled as a linear function of independent variables. Model parameters are estimated such that they minimised the sum of the absolute differences between observed and predicted values. *P* values were generated using the wild bootstrap method⁵³ to estimate variance.

A linear quantile (median) regression for two variables—a response variable (*y*) and a predictor variable (*x*)—is the following: median (*y*) = $\beta_0 + \beta_1 x$ where β_0 is the intercept (value when *y* = 0) and β_1 is the slope (change in median of *y* for a unit change in *x*). Together, these parameters describe the association between *y* and *x*, where *x* is a predictor of *y*. In the case of multiple predictor variables, each one is added to the regression equation and so the equation becomes median (*y*) = $\beta_0 + \beta_1 x_1 + \beta_2 x_2$ and now the slope β_1 is interpreted as the median change in *x*₁ after adjusting for *x*₂. This can be likened to a laboratory experiment where the specific effect of one variable on another is isolated by holding all other relevant variables constant.

Following statistical analysis of the taxonomic classifications, we estimated FDR values using the Benjamini–Hochberg method⁵⁴ to control for multiple testing. The exception to this were analyses at the genus level where we estimated the proportion of true null hypotheses with the *Q*-value function unless the estimated π_0 was less than or equal to zero⁵⁵.

Statistical analysis of the NMR data was performed using diverse software packages: PCA and PLS-DA analysis was performed in SIMCA-P+ (Umetrics); permutation testing was performed in R and PPCCA was performed in R using the MetabolAnalyze package. The NMR data was Pareto-scaled before data analysis. Assignment of the spectral peaks (Supplementary Table 9) was performed using in-house libraries, statistical correlation analysis and two-dimensional NMR spectra (TOCSY and COSY).

42. Harrington, J. *et al.* Sociodemographic, health and lifestyle predictors of poor diets. *Public Health Nutr.* **14**, 2166–2175 (2011).
43. McCance, R. A. & Widdowson, E. M. *The composition of foods* 6th edn (Royal Soc. Chemistry, 2002).
44. Claesson, M. J. *et al.* Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PLoS ONE* **4**, e6669 (2009).
45. Lilburn, T. G. & Garrity, G. M. Exploring prokaryotic taxonomy. *Int. J. Syst. Evol. Microbiol.* **54**, 7–13 (2004).
46. Caporaso, J. G. *et al.* PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**, 266–267 (2010).
47. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
48. Hamady, M., Lozupone, C. & Knight, R. Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J.* **4**, 17–27 (2010).
49. Namiki, T., Hachiya, T., Tanaka, H. & Sakakibara, Y. in *ACM Conference on Bioinformatics Computational Biology and Biomedicine* (Association for Computing Machinery, 2011).
50. Chevreux, B. *et al.* Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res.* **14**, 1147–1159 (2004).
51. Noguchi, H., Park, J. & Takagi, T. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res.* **34**, 5623–5630 (2006).
52. Koenker, R. & Basset, G. Regression quantiles. *Econometrica* **46**, 33–50 (1978).
53. Feng, X. D., He, X. M. & Hu, J. H. Wild bootstrap for quantile regression. *Biometrika* **98**, 995–999 (2011).
54. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).
55. Dabney, A., Storey, J. D. & Warnes, G. R. *qvalue: Q-value estimation for false discovery rate control*; R package version 1.24.20 (2010).

Quantum teleportation and entanglement distribution over 100-kilometre free-space channels

Juan Yin^{1*}, Ji-Gang Ren^{1*}, He Lu^{1*}, Yuan Cao¹, Hai-Lin Yong¹, Yu-Ping Wu¹, Chang Liu¹, Sheng-Kai Liao¹, Fei Zhou¹, Yan Jiang¹, Xin-Dong Cai¹, Ping Xu¹, Ge-Sheng Pan¹, Jian-Jun Jia², Yong-Mei Huang³, Hao Yin¹, Jian-Yu Wang², Yu-Ao Chen¹, Cheng-Zhi Peng¹ & Jian-Wei Pan¹

Transferring an unknown quantum state over arbitrary distances is essential for large-scale quantum communication and distributed quantum networks. It can be achieved with the help of long-distance quantum teleportation^{1,2} and entanglement distribution. The latter is also important for fundamental tests of the laws of quantum mechanics^{3,4}. Although quantum teleportation^{5,6} and entanglement distribution^{7–9} over moderate distances have been realized using optical fibre links, the huge photon loss and decoherence in fibres necessitate the use of quantum repeaters¹⁰ for larger distances. However, the practical realization of quantum repeaters remains experimentally challenging¹¹. Free-space channels, first used for quantum key distribution^{12,13}, offer a more promising approach because photon loss and decoherence are almost negligible in the atmosphere. Furthermore, by using satellites, ultra-long-distance quantum communication and tests of quantum foundations could be achieved on a global scale. Previous experiments have achieved free-space distribution of entangled photon pairs over distances of 600 metres (ref. 14) and 13 kilometres (ref. 15), and transfer of triggered single photons over a 144-kilometre one-link free-space channel¹⁶. Most recently, following a modified scheme¹⁷, free-space quantum teleportation over 16 kilometres was demonstrated¹⁸ with a single pair of entangled photons. Here we report quantum teleportation of independent qubits over a 97-kilometre one-link free-space channel with multi-photon entanglement. An average fidelity of 80.4 ± 0.9 per cent is achieved for six distinct states. Furthermore, we demonstrate entanglement distribution over a two-link channel, in which the entangled photons are separated by 101.8 kilometres. Violation of the Clauser–Horne–Shimony–Holt inequality⁴ is observed without the locality loophole. Besides being of fundamental interest, our results represent an important step towards a global quantum network. Moreover, the high-frequency and high-accuracy acquiring, pointing and tracking technique developed in our experiment can be directly used for future satellite-based quantum communication and large-scale tests of quantum foundations.

Following the original quantum teleportation scheme^{1,2}, Alice and Bob share an entangled photon pair distributed by Charlie. An unknown state can be teleported from Alice to Bob by performing a joint Bell-state measurement on the two photons with Alice (see Supplementary Fig. 1a for details). Experimentally, we start with an ultra-bright entangled photon source¹⁹ based on type-II spontaneous parametric down-conversion²⁰. On Charlie's side (located at Gangcha next to Qinghai Lake; latitude $37^\circ 16' 42.41''$ N, longitude $99^\circ 52' 59.88''$ E; altitude 3,262 m; Fig. 1a), an entangled photon pair 2 and 3 in state $|\Phi^+\rangle_{23} = (|HH\rangle_{23} + |VV\rangle_{23})/\sqrt{2}$ is created by Charlie and then distributed to Alice and Bob, where H (V) represents the horizontal (vertical) polarization of the photonic state (see Fig. 1b and Methods). An average twofold coincidence rate of $4.4 \times 10^5 \text{ s}^{-1}$ for the entangled photon source was observed.

To prepare the unknown state to be teleported, Alice uses an ultra-violet laser to pump a collinear β -barium borate (BBO) crystal, which emits photon pairs in $|HV\rangle_{14}$ along the pumping direction (see Fig. 1c). After filtering out the pumping laser, a polarized beam splitter (PBS) splits the photon pair. A twofold coincidence rate of $6.5 \times 10^5 \text{ s}^{-1}$ was observed. A half-wave plate (HWP) and a quarter-wave plate (QWP) are applied to photon 1 to prepare the initial state $|\chi\rangle_1 = \alpha|H\rangle + \beta|V\rangle$, when triggered by photon 4. Alice then performs a joint Bell-state measurement on photons 1 and 2 by interfering them on a PBS and performing polarization analysis on the two outputs. The subsequent coincidence measurements can identify the $|\Phi^\pm\rangle$ Bell states in our experiment. In the joint Bell-state measurement, the observed visibility of interference on the PBS was 0.6. Finally, we observed a fourfold coincidence of $2 \times 10^3 \text{ s}^{-1}$ locally. Such a brightness supports successful quantum teleportation over a high channel loss, which can be greater than 50 dB (see Supplementary Fig. 1b and c).

Charlie sends photon 3 (by means of a compact transmitting system) through a 97-km free-space channel to Bob (Fig. 1b). A 127-mm $f/7.5$ (that is, aperture 127 mm, focal length 952.5 mm) extra-low dispersion apochromatic refractor telescope is used as an optical transmitting antenna. For near-diffraction-limited far-field divergence angles, we have designed our systems to reduce chromatic and spherical aberrations substantially. The divergence angle of our compact quantum transmitter is about $20 \mu\text{rad}$.

As shown in Fig. 1d, on the other side of Qinghai Lake (Guanjing; $36^\circ 32' 43.31''$ N, $100^\circ 28' 9.81''$ E; altitude 3,682 m), Bob receives photon 3 in a 400-mm-diameter off-axis reflecting telescope. An integrated measurement system, consisting of an HWP, a QWP and a PBS, is assembled at the telescope's exit for state analysis. Passing through two band-pass filters (full-width at half-maximum bandwidth $\Delta_{\text{FWHM}} = 80 \text{ nm}$) and one narrow-band interference filter ($\Delta_{\text{FWHM}} = 10 \text{ nm}$) used to reduce background noise (IF, shown in Fig. 1d), the photons are coupled in multi-mode fibres and then detected by the single-photon counting modules (SPCMs) with ultra-low dark counts ($<20 \text{ s}^{-1}$). The noise that we observed, including the dark counts and ambient counts, was in total about 160 s^{-1} to 300 s^{-1} , depending on the position of the Moon. On average, we obtained about 200 s^{-1} .

In addition, we assembled an acquiring, pointing and tracking (APT) system to account for effects due to ground settlement, mechanical deformation, atmospheric turbulence and so on. As shown in Fig. 1b, when the optical link is established for the first time, Charlie acquires the signal by global positioning system coordinates and a light guide. At the same time, he switches on the beacon laser (532 nm) pointing to the receiver, Bob. Bob then achieves acquisition and fires another beacon laser (671 nm) pointing back to Charlie.

The tracking system is composed of cascaded closed-loop control systems (the blue and green arrows in Fig. 1b and d). On Charlie's side,

¹Shanghai Branch, National Laboratory for Physical Sciences at Microscale and Department of Modern Physics, University of Science and Technology of China, Shanghai 201315, China. ²Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China. ³The Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu 610209, China.

*These authors contributed equally to this work.

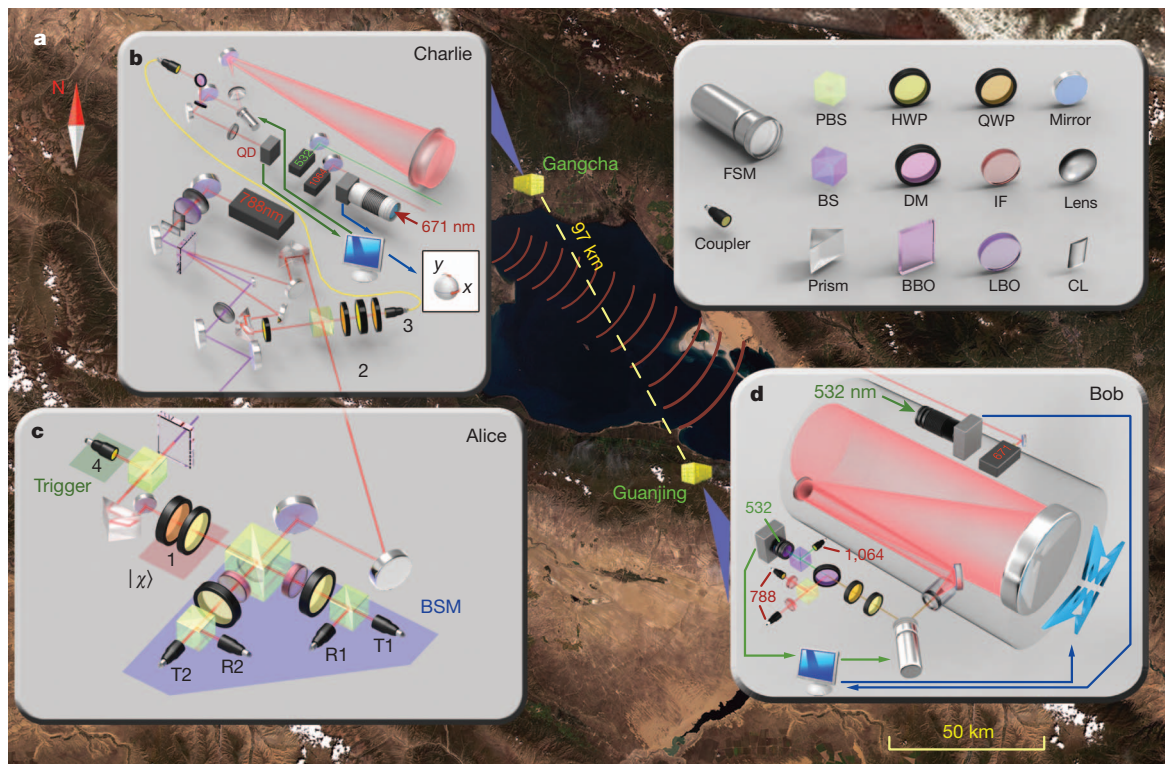


Figure 1 | Bird's-eye view and schematic diagram for free-space quantum teleportation. **a**, Satellite image of experiment site. **b**, Entanglement generation and distribution on Charlie's side. Using a pulsed 788-nm laser, an entangled photon pair (2 and 3) is created by Charlie (see Methods). Photon 2 is sent to Alice for a Bell-state measurement (BSM), and photon 3 is guided to a refractor telescope and sent to Bob. Coaxial with the telescope, there is a green laser (532 nm, 200 mW, 1.5 mrad) for system tracking and a pulsed infrared laser (1,064 nm, 10 kHz, 50 mW, 200 μ rad) for synchronization (Supplementary Information). Lasers are shown as black boxes labelled with emission wavelength. x and y denote the azimuth and elevation axis of the rotatable platform for the transmitting telescope. **c**, Initial state preparation and BSM on

Alice's side. Under a trigger of photon 4, photon 1 is prepared in the initial state $|\chi\rangle$. A coincidence between detectors T1 and T2 (R2) or R1 and R2 (T2) indicates the incident state of $|\Phi^+\rangle$ ($|\Phi^-\rangle$). **d**, Receiving system and polarization analysis on Bob's side. The high-power beacon laser (671 nm, 2 W, 200 μ rad) is used for system tracking. The DM is used to separate the signal photon (788 nm) from the tracking light (532 nm) and the synchronization light (1,064 nm), which are then detected by corresponding detectors. The blue and green arrows on both sides indicate the coarse and fine tracking systems, respectively. Key at top right defines symbols used for the set-up. Image in **a** was obtained by the Chinese environmental satellites, and is used with permission.

the beacon laser from the receiver, Bob, is detected by a wide-angle camera. Using a feedback loop, coarse alignment of the entire optical system is achieved by a platform rotatable in both azimuth and elevation (blue arrows in Fig. 1b). Similarly, the fine tracking indicated by the green arrows is achieved by a fast steering mirror (FSM) driven by piezo ceramics with the feedback from the four-quadrant detector (QD). Furthermore, the fine tracking system shares the same optical path as the quantum channel and is later separated by a dichroic mirror (DM). A high tracking accuracy can be obtained. The closed-loop bandwidth of the fine tracking is more than 150 Hz (see Supplementary Figs 2 and 3 for a detailed description for the APT system), which is sufficient to overcome most of the atmospheric turbulence²¹. Finally, with this system design the tracking accuracy is better than 3.5 μ rad over the 97-km free-space link.

There is also coarse and fine tracking on Bob's side, by means of closed-loop control of the telescope's own motor and FSM (Fig. 1d). Because the main purpose of the tracking at the receiver is to reduce the low-frequency shaking due to ground settlement and passing vehicles, the tracking bandwidth is about 10 Hz. The APT system is designed for tracking an arbitrarily moving object, and can be directly used for a satellite-based quantum communication experiment. In experiments between fixed locations, the first two steps, acquiring and pointing, do not need to be done every day.

After debugging the entire system, the channel loss of the 97-km horizontal atmospheric transmission at near ground level was measured to be between 35 and 53 dB, of which 8 dB was due to the imperfect optics and finite collection efficiency, and 8–12 dB was due

to atmospheric loss. The geometric attenuation caused by the beam spreading wider than the aperture of the receiver telescope was between 19 and 33 dB, corresponding to a far-field spot size of between 3.5 and 17.9 m, depending on weather conditions. With a tracking accuracy of 3.5 μ rad (a pointing error of 0.34 m at the receiver), we had stable count rates for single photons. The average channel attenuation was about 44 dB, and the time synchronization accuracy was better than 1 ns (see Supplementary Fig. 5 for details). Finally, we obtained 1,171 coincidences during an effective time of 14,400 s. Six distinct polarization states, namely, $|H\rangle$, $|V\rangle$, $|\pm\rangle = (|H\rangle \pm |V\rangle)/\sqrt{2}$, $|R\rangle = (|H\rangle + i|V\rangle)/\sqrt{2}$ and $|L\rangle = (|H\rangle - i|V\rangle)/\sqrt{2}$ were teleported. The experimental fidelities for the six teleported states range from 76% to 89%, all well beyond the classical limit²² of 2/3, with an overall average fidelity of 80% (Table 1).

In the present teleportation experiment, Alice and Charlie are close to each other. A more common situation would be that Alice is also far away from Charlie. In this case, distribution of entanglement between

Table 1 | Fidelity of quantum teleportation over 97 km

State	Fidelity
H	0.814 ± 0.031
V	0.886 ± 0.024
$+$	0.773 ± 0.031
$-$	0.781 ± 0.031
R	0.808 ± 0.026
L	0.760 ± 0.027

The data were accumulated for 14,400 s. Errors shown are statistical errors, ± 1 s.d.

Alice and Bob is a prerequisite for quantum teleportation. A feasible solution is to distribute the entanglement by Charlie via a two-link channel. To demonstrate the two-link entanglement distribution, we move the entanglement source close to the middle of the free-space channel, an island in the middle of Qinghai Lake (Haixin, $36^{\circ} 51' 38.75''$ N, $100^{\circ} 8' 15.22''$ E, Fig. 2a). In order to show a two-link entanglement distribution between two sites, which cannot see each other directly, Bob moves his receiving platform to a different position, provided by a local Tibetan family (Gonghe; $36^{\circ} 32' 20.66''$ N, $100^{\circ} 33' 45.38''$ E), that is next to Guanjing (Fig. 2a). Charlie first prepares the entangled photon pairs in the state $|\Phi^+\rangle$, which are then sent to Alice and Bob via two telescopes each mounted on a two-dimensional rotatable platform (Fig. 2b–d). The distances between Charlie and the two receivers are 51.2 km (Alice) and 52.2 km (Bob), and the distance between Alice and Bob is 101.8 km.

Whereas the same APT system as in the teleportation experiment is used between Bob and Charlie, the APT system used between Alice and Charlie is slightly modified (see Supplementary Fig. 4 for details). Entangled photons are then collected by telescopes on both sides. In contrast to the 400-mm off-axis reflecting telescope used on Bob's side, Alice uses a 600-mm Cassegrain telescope to collect the photons. To confirm the successful entanglement distribution between the two receivers, we measure the S parameter in the Clauser–Horne–Shimony–Holt (CHSH)-type Bell's inequality¹⁷ defined as $S = |E(\varphi_A, \varphi_B) - E(\varphi_A, \varphi'_B) - E(\varphi'_A, \varphi_B) + E(\varphi'_A, \varphi'_B)|$, where $E(\varphi_A, \varphi_B)$

is the correlation function, and φ_A and φ'_A (φ_B and φ'_B) the measurement settings of the photon in Alice's (Bob's) site. The settings are randomly selected among $(0, \pi/8)$, $(0, 3\pi/8)$, $(\pi/4, \pi/8)$ and $(\pi/4, 3\pi/8)$ by two fast electro-optical modulators (EOMs) and their logical circuits controlled by two quantum random number generators (not shown). The optical axes of the EOMs are set at 22.5° such that the EOMs act as 22.5° HWPs when half-wave voltages are applied and act as absent wave plates when zero-wave voltages are imposed. Quantum random number generators are used to produce the random digital series between zero-wave and half-wave voltages. Together with the HWP (0° at Alice's side and 11.25° at Bob's side) and the quantum random number generator, the EOM randomly switches between the two desired measurement bases: 0 and $\pi/4$ for Alice and $\pi/8$ and $3\pi/8$ for Bob.

Finally, we obtained 208 coincidences during an effective time of 32,000 s. By comparison with the counts of our entanglement source, we found that the channel attenuation varied from 66 dB to 85 dB with an average value of 79.5 dB. For 20-cm-aperture satellite optics at an orbit height of 600 km and 1-m-aperture receiving optics, the total loss for a two-downlink channel between a satellite and two ground stations is typically about 75 dB. The measured correlation functions (shown in Fig. 3) resulted in $S = 2.51 \pm 0.21$, which violates Bell's inequality by 2.4 standard deviations. This shows the feasibility of quantum teleportation with the modified scheme¹⁷. Multi-photon, multi-link quantum teleportation, however, remains experimentally



Figure 2 | Illustration of the experimental set-up for entanglement distribution. **a**, Satellite image of experiment site. CMOS, complementary metal–oxide–semiconductor camera. **b**, Alice collects the photon sent by Charlie using a 600-mm Cassegrain telescope. **c**, At the receiver, Bob, the photon beams are collected by a 400-mm off-axis reflecting telescope. **d**, An entangled photon pair ($|\Phi^+\rangle$) is created by Charlie at the centre island of

Qinghai Lake (Haixin) and then distributed to Alice and Bob. After receiving the photons, Alice and Bob guide the photons to the detection module by an optical system to measure the S parameter in the CHSH-type Bell's inequality. Image in **a** was obtained by the Chinese environmental satellites, and is used with permission.

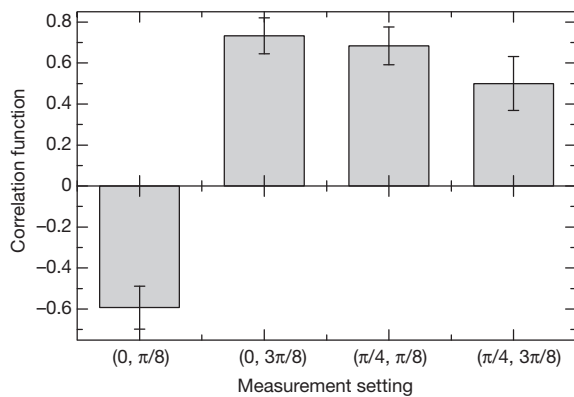


Figure 3 | Correlation functions of a CHSH-type Bell's inequality for entanglement distribution. The measurement setting (φ_A, φ_B) represents the measured polarization bases of photons by Alice and Bob, respectively. Error bars, statistical errors (± 1 s.d.).

challenging. The main difficulty is the spatial mode mismatching at the Bell-state measurement caused by atmospheric turbulence, which requires increasing the tracking bandwidth to more than one kilohertz: this requires the use of new technologies—such as an adaptive optics system, and photon collection with single-mode fibres.

In addition, our experiment closed the locality loophole. The entangled photon pairs were distributed along two opposite directions to Alice and Bob: these parties are separated by 101.8 km, a distance that takes 340 μ s for light to travel, and the path difference of the two links is 1 km, which results in a 3- μ s delay between the two measurement events. Thus, the two measurement events on Alice's and Bob's sites are space-like separated. Furthermore, the two receivers used fast EOMs to switch between the two possible polarization bases. The two EOMs were controlled by two independent quantum random number generators, each of which generates a random number every 20 μ s (less than 340 μ s). Thus the measurement-setting choices are also space-like separated. Hence, the locality loophole is closed.

In this work, we experimentally realized free-space quantum teleportation for independent qubits over a 35–53-dB-loss one-link channel. In comparison with previous multi-photon experiments^{8,9}, we have enhanced the transmission distance by two orders of magnitude to 97 km. Furthermore, we demonstrated the distribution of entangled photon pairs over a two-link free-space channel to two receivers separated by more than 100 km. In contrast to previous long-distance free-space experiments with entangled photon pairs using only one-link channels^{15,23}, our two-link experiment requires tracking and synchronization between three different locations. Our two-link experiment—most comparable with satellite-to-ground quantum entanglement distribution—has achieved a distance between two receivers that is an order of magnitude larger than in previous experiments. Our results show that even with a high-loss ground-to-satellite uplink channel, or satellite-to-ground two-downlink channel, quantum teleportation and entanglement distribution can be realized. Furthermore, our APT system can be used to track an arbitrarily moving object with high frequency and high accuracy, which is essential for future satellite-based ultra-long-distance quantum communication. We believe our experiment will help fundamental tests of the laws of quantum mechanics on a global scale to be achieved.

METHODS SUMMARY

Entangled photon source. As shown in Fig. 1b, a femtosecond ultraviolet laser (394 nm, 1.3 W) is created by frequency doubling a pulsed laser (central wavelength of 788 nm with a pulse duration of 130 fs, a repetition rate of 76 MHz and an average power of about 3 W) with a LiB₃O₅ (LBO) crystal. After optimizing the beam profile with two cylindrical lenses (CL), the ultraviolet laser then pumps a type-II β -barium (BBO) crystal, creating a pair of polarization entangled photons in the state

$|\Psi\rangle = (|H_o V_e\rangle + |V_e H_o\rangle)/\sqrt{2}$ with temporal and polarization information both entangled¹⁹, where ordinary (o) ray and extraordinary (e) ray indicate the polarizations with respect to the pump. With an interferometric Bell-state synthesizer^{19,24}, we disentangle the temporal from the polarization information by guiding photons of different bandwidths through separate paths, resulting in the desired entangled state $|\Phi^+\rangle_{23} = (|HH\rangle_{23} + |VV\rangle_{23})/\sqrt{2}$. With a 3-nm filter in the e-ray path of photon 2 and an 8-nm filter in the o-ray path of photon 3, we observed an average twofold coincidence rate of 4.4×10^5 s⁻¹ with a visibility of about 91% in the $|H\rangle|V\rangle$ basis and 90% in the $|+\rangle|-\rangle$ basis. The generation rate of the entangled photons was about 0.1 pairs per pulse, and the overall detection efficiency was 23.6% locally.

Received 20 January; accepted 15 June 2012.

- Bennett, C. H. *et al.* Teleporting an unknown quantum state via dual classic and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **70**, 1895–1899 (1993).
- Bouwmeester, D. *et al.* Experimental quantum teleportation. *Nature* **390**, 575–579 (1997).
- Bell, J. S. On the Einstein-Podolsky-Rosen paradox. *Physics* **1**, 195–200 (1964).
- Clauser, J. F., Home, M. A., Shimony, A. & Holt, R. A. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **23**, 880–884 (1969).
- Marcikic, I., de Riedmatten, H., Tittel, W., Zbinden, H. & Gisin, N. Long-distance teleportation of qubits at telecommunication wavelengths. *Nature* **421**, 509–513 (2003).
- Ursin, R. *et al.* Quantum teleportation across the Danube. *Nature* **430**, 849 (2004).
- Hübel, H. *et al.* High-fidelity transmission of polarization encoded qubits from an entangled source over 100 km of fiber. *Opt. Express* **15**, 7853–7862 (2007).
- Zhang, Q. *et al.* Distribution of time-energy entanglement over 100 km fiber using superconducting single photon detectors. *Opt. Express* **16**, 5776–5781 (2008).
- Dynes, J. F. *et al.* Efficient entanglement distribution over 200 kilometers. *Opt. Express* **17**, 11440–11449 (2009).
- Briegleb, H. J., Dür, W., Cirac, J. I. & Zoller, P. Quantum repeaters: the role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81**, 5932–5935 (1998).
- Yuan, Z.-S. *et al.* Experimental demonstration of a BDCZ quantum repeater node. *Nature* **454**, 1098–1101 (2008).
- Hughes, R. J., Nordholt, J. E., Derkacs, D. & Peterson, C. G. Practical free-space quantum key distribution over 10 km in daylight and at night. *N. J. Phys.* **4**, 43 (2002).
- Kurtsiefer, C. *et al.* Quantum cryptography: a step towards global key distribution. *Nature* **419**, 450 (2002).
- Aspelmeyer, M. *et al.* Long-distance free-space distribution of quantum entanglement. *Science* **301**, 621–623 (2003).
- Peng, C.-Z. *et al.* Experimental free-space distribution of entangled photon pairs over 13 km: towards satellite-based global quantum communication. *Phys. Rev. Lett.* **94**, 150501 (2005).
- Fedrizzi, A. *et al.* High-fidelity transmission of entanglement over a high-loss free-space channel. *Nature Phys.* **5**, 389–392 (2009).
- Boschi, D., Branca, S., Martini, F. D., Hardy, L. & Popescu, S. Experimental realization of teleporting an unknown pure quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **80**, 1121–1125 (1998).
- Jin, X.-M. *et al.* Experimental free-space quantum teleportation. *Nature Photon.* **4**, 376–381 (2010).
- Yao, X.-C. *et al.* Observation of eight-photon entanglement. *Nature Photon.* **6**, 225–228 (2012).
- Kwiat, P. G. *et al.* New high-intensity source of polarization-entangled photon pairs. *Phys. Rev. Lett.* **75**, 4337–4341 (1995).
- Strohbehn, J. W. *Laser Beam Propagation in the Atmosphere* (Springer, 1978).
- Popescu, S. Teleportation versus Bell's inequalities. What is nonlocality? *Phys. Rev. Lett.* **72**, 797–799 (1994).
- Fedrizzi, A. *et al.* High-fidelity transmission of entanglement over a high-loss free-space channel. *Nature Phys.* **5**, 389–392 (2009).
- Kim, Y.-H., Kulik, S. P., Chekhova, M. V., Grice, W. P. & Shih, Y. Experimental entanglement concentration and universal Bell-state synthesizer. *Phys. Rev. A* **67**, 010301 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to the staff of the Qinghai Lake National Natural Reserve Utilization Administration Bureau, especially Y.-B. He and Z. Xing, for their support during the experiment. We thank B. Zhao and Y.-J. Deng for discussions, and T. Li for proofreading the manuscript before submission. This work was supported by Chinese Academy of Sciences, National Natural Science Foundation of China and the National Fundamental Research Program (under grant no. 2011CB921300).

Author Contributions All authors contributed extensively to the work presented in this paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Y.-A.C. (yuaochen@ustc.edu.cn) or C.-Z.P. (pcz@ustc.edu.cn) or J.-W.P. (pan@ustc.edu.cn).

A III–V nanowire channel on silicon for high-performance vertical transistors

Katsuhiko Tomioka^{1,2}, Masatoshi Yoshimura¹ & Takashi Fukui¹

Silicon transistors are expected to have new gate architectures, channel materials and switching mechanisms in ten years' time^{1–4}. The trend in transistor scaling has already led to a change in gate structure from two dimensions to three, used in fin field-effect transistors, to avoid problems inherent in miniaturization such as high off-state leakage current and the short-channel effect. At present, planar and fin architectures using III–V materials, specifically InGaAs, are being explored as alternative fast channels on silicon^{5–9} because of their high electron mobility and high-quality interface with gate dielectrics¹⁰. The idea of surrounding-gate transistors¹¹, in which the gate is wrapped around a nanowire channel to provide the best possible electrostatic gate control, using InGaAs channels on silicon, however, has been less well investigated^{12,13} because of difficulties in integrating free-standing InGaAs nanostructures on silicon. Here we report the position-controlled growth of vertical InGaAs nanowires on silicon without any buffering technique and demonstrate surrounding-gate transistors using InGaAs nanowires and InGaAs/InP/InAlAs/InGaAs core-multishell nanowires as channels. Surrounding-gate transistors using core-multishell nanowire channels with a six-sided, high-electron-mobility transistor structure greatly enhance the on-state current and transconductance while keeping good gate controllability. These devices provide a route to making vertically oriented transistors for the next generation of field-effect transistors and may be useful as building blocks for wireless networks on silicon platforms.

In_xGa_{1–x}As has attracted attention as an alternative fast channel on silicon for future n-type field-effect transistors (FETs) because of its higher electron mobility compared with that of bulk silicon and its small electron effective mass^{13–16}. The Schottky barrier height at the insulator–In_xGa_{1–x}As interface can be adjusted by changing the In content¹⁰. Using In_{0.7}Ga_{0.3}As fin FETs on Si, it has recently been shown that device performance can be improved by exploiting the three-dimensional architecture and III–V materials^{8,9}. As yet, however, there has been less investigation of vertical nanoscale transistors, such as surrounding-gate transistors (SGTs) or gate-all-around FETs using the In_{0.7}Ga_{0.3}As channel on Si (ref. 13). This is because new epitaxial techniques for integrating one-dimensional InGaAs nanostructures on Si substrates must be developed to deal with crystallographic defects (threading dislocations, misfit dislocations, antiphase defects and so on) that form as a result of mismatched lattice constants, thermal expansion coefficients and polarities. Several methods using buffering techniques^{8,17} have been investigated, but none of them can completely avoid the formation of defects.

Recent progress in selective-area metal–organic vapour phase epitaxy has allowed position-controlled integration of vertical III–V nanowires on Si substrates, regardless of mismatches in terms of the lattice constant, thermal expansion coefficient and polarity^{18,19}. We further developed this bottom-up technique for aligning vertical In_{0.7}Ga_{0.3}As nanowires on Si. The important step in aligning vertical In_{0.7}Ga_{0.3}As nanowires on Si(111) is that of forming a (111)B polar surface on a non-polar Si(111) surface by changing from complex Si

surface reconstructions to the 1 × 1 structure¹⁸. This can be done by annealing the Si(111) substrate at 900 °C in H₂ and then treating it in an AsH₃ atmosphere at 670 °C. During this treatment, the outermost Si atoms are replaced by As atoms in the AsH₃ atmosphere¹⁸. Flow-rate modulation epitaxy in which group-III and group-V gas source materials are alternately supplied is used to terminate the remaining dangling bonds on Si(111) with group-III atoms (Supplementary Fig. 1). The Si(111) surface terminated with these atoms is isostructural with the (111)B polar surface.

Figure 1a shows typical InGaAs nanowires on Si(111) grown by selective-area metal–organic vapour phase epitaxy (Supplementary Fig. 2). Vertically aligned InGaAs nanowires with {110} side walls and (111)B top surfaces were grown on the Si(111) substrate. The electron transport channel is in the [111] direction in the {110} planes. The transmission electron microscopy image in Fig. 1b reveals that the {110} side walls had atomically flat surfaces and the crystal structure was that of zincblende with rotational twins. The average diameter of the InGaAs nanowires was 90 nm, and the average height was 760 nm. The energy-dispersive X-ray spectroscopy (EDX) line-scan profile in Fig. 1c indicated that the In content of the InGaAs nanowire was 70 ± 2%. The In content remained constant from the bottom to the top of the nanowire. Monosilane (SiH₄) gas was used for n-type doping. The carrier concentration (*n*) for the n-InGaAs nanowire, as evaluated from both a four-terminal current–voltage measurement²⁰ and the EDX analysis, was approximately 1 × 10¹⁸ cm^{–3}. Figure 1d, e shows the interface of an InGaAs nanowire epitaxially grown on the Si substrate. The heterointerface (Fig. 1d) was apparently free of threading dislocations and antiphase defects, but had periodical misfit dislocations (Fig. 1f). The average period of the misfit dislocation was 38 ± 3 Å. These dislocation networks were formed only at the heterointerface (within three monolayers; see Supplementary Fig. 3).

To characterize the transistor performance of the InGaAs nanowires on Si, we made vertical SGTs (Fig. 2a). We used a structure of stacked Si-doped (*n* = 1 × 10¹⁸ cm^{–3}) and undoped (*n* = 5 × 10¹⁶ cm^{–3}) InGaAs nanowires with diameters of 60 nm. The heights of the undoped and Si-doped nanowires were 180 nm and 1.2 μm, respectively. The fabrication processes included deposition of a low-*k* polymer (benzocyclobutene), and slight etch back using reactive-ion etching (Supplementary Fig. 4). We used Hf_{0.8}Al_{0.2}O as gate oxide. The gate length (*L_G*) was 200 nm, and the gate–drain distance (*L_{G–D}*) was 50 nm. Fabricated devices include ten nanowires connected with each other in parallel to a contact pad. The device performance was evaluated at room temperature in the dark.

Figure 2b depicts the capacitance/gate voltage (*C*–*V_G*) curves of a single InGaAs nanowire SGT (the nanowire diameter was 90 nm) with an effective oxide thickness (EOT) of 2.75 nm. In this measurement, the drain metal was grounded to avoid the parasitic capacitance of the benzocyclobutene and SiO₂ layers on the Si, and the device comprises 250 nanowires in parallel. The capacitance shown in Fig. 2b is that for a single wire (the measured capacitance was divided by the number of nanowires). We note that in the accumulation region of the *C*–*V_G*

¹Graduate School of Information Science and Technology, and Research Center for Integrated Quantum Electronics (RCIQE), Hokkaido University, Kita 13 Nishi 8, Sapporo 060-8628, Japan. ²Japan Science and Technology Agency—PRESTO, Kawaguchi, Saitama 332-0012, Japan.

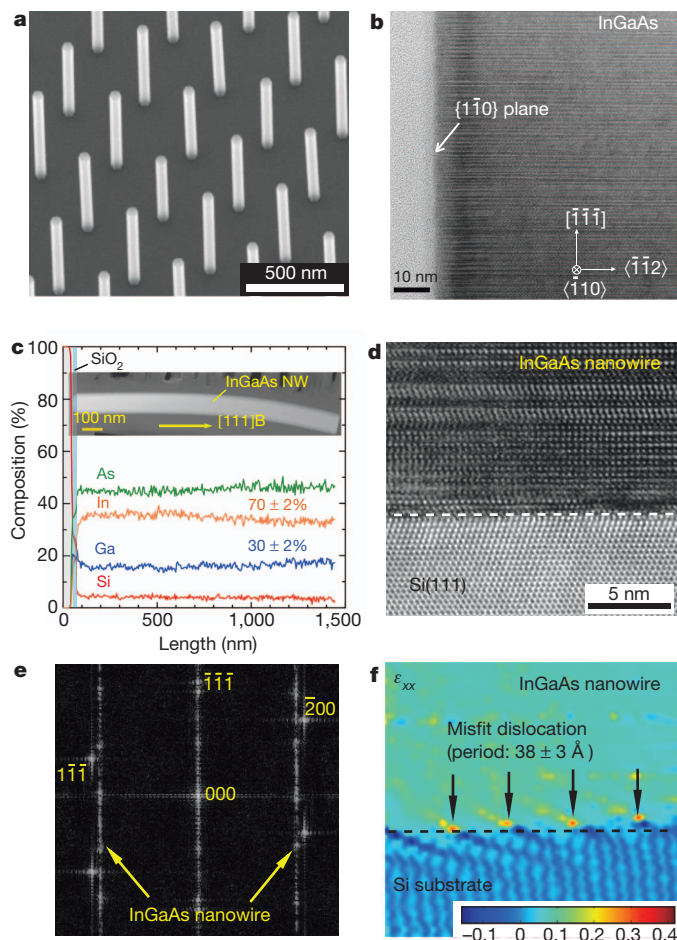


Figure 1 | Selective-area growth of InGaAs nanowires on Si(111). **a**, Typical scanning electron microscope (SEM) image showing selective-area growth of InGaAs nanowires on Si: vertically aligned nanowires were grown on a Si(111) substrate. The surface between the nanowires is covered with a SiO₂ film. The nanowires are 90 nm in diameter and 760 nm in height. They are surrounded by six-sided {110} planes and a (111)B top surface. The electron transport is along the [111] direction. **b**, High-resolution transmission electron microscope (TEM) image of an InGaAs nanowire. The electron beam is incident in the [110] directions. The crystal structure is that of zincblende with rotational twins. **c**, Atomic content profiles for Si (red), Ga (blue), As (green) and In (yellow) evaluated along the wire by EDX line-scan profiling. The inset is a high-angle, annular dark-field scanning TEM (HAADF-STEM) image of an InGaAs nanowire with a diameter of 130 nm and a height of 1.4 μm. The In and Ga contents of the InGaAs nanowire were approximately 70 ± 2% and 30 ± 2%, respectively. **d**, High-resolution TEM image of near the heterointerface (dashed line) between the InGaAs nanowire and Si. **e**, Fast Fourier transformation image of **d**, showing diffraction spots of Si and InGaAs. The overlap indicates that the InGaAs nanowire is epitaxially grown on Si. **f**, Strain mapping (ϵ_{xx}) estimated from a filtered version of **d**.

curves in Fig. 2b, the frequency dispersion between 10 kHz and 1 MHz was less than 4%, and that in this frequency range the flat-band voltage was no longer shifted. The $C-V_G$ curves had no noticeable hysteresis, as revealed by bidirectional sweeps of V_G , and the gate leakage current obtained for various EOTs was small: 10^{-3} – 10^{-5} A cm⁻² (Supplementary Fig. 5). These data indicate that the Hf_{0.8}Al_{0.2}O–In_{0.7}Ga_{0.3}As interface was of good quality, with few interface states.

Figure 2c, d show the transfer and output characteristics of the InGaAs nanowire SGT on Si. The drain currents are those for a single wire (calculated as for Fig. 2b) and are normalized by the outer perimeter of the gate, that is, channel width. The switching properties showed an n-type enhancement mode with a threshold voltage (V_T) of 0.18 V. The source–drain current (I_D) was modulated by V_G with a minimum subthreshold slope ($SS = dV_G/d[\log(I_D)]$) of 85 mV per

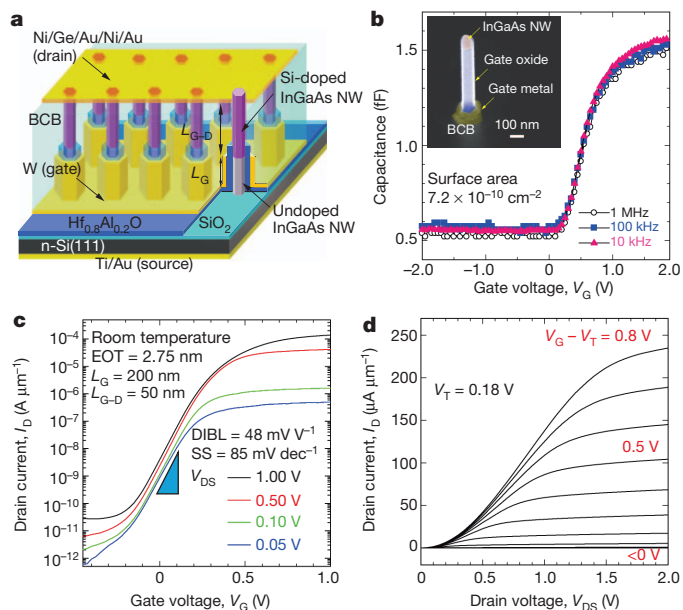


Figure 2 | Vertical InGaAs nanowire channel SGT on Si. **a**, Structure of an SGT with ten nanowires (NWs) connected in parallel with the drain metal. Each nanowire comprises Si-doped InGaAs ($n = 1 \times 10^{18}$ cm⁻³) stacked on undoped InGaAs ($n = 5 \times 10^{16}$ cm⁻³) and has a diameter of 60 nm. The heights of the undoped and Si-doped InGaAs nanowires are 180 nm and 1.2 μm, respectively. Each nanowire is wrapped with 14-nm-thick Hf_{0.8}Al_{0.2}O gate oxide and tungsten (W) gate metal. The drain and source metals are respectively Ni/Ge/Au/Ni/Au and Ti/Au. The gate length and gate–drain distance are 200 nm and 50 nm, respectively. **b**, Capacitance/gate voltage curves of the SGT (nanowire diameter, 90 nm). The sample has 250 nanowires aligned in parallel. The capacitance indicated is that for a single wire. Inset is a representative SEM image showing the SGT structure. In this measurement, the drain metal is grounded to avoid the parasitic capacitance of the SGT and benzocyclobutene (BCB). **c**, Transfer characteristics of an SGT with InGaAs nanowires (nanowire diameter, 60 nm). The drain current indicated is that for one wire and is normalized by the outer perimeter of the gate (180 nm). The SS and DIBL are 85 mV per decade and 48 mV V⁻¹, respectively. **d**, Output characteristics of an SGT with InGaAs nanowire channels. The gate voltage (bias) is changed from –0.4 to 1.0 V in steps of 0.1 V. The threshold voltage is estimated to be 0.18 V using linear extrapolation of the output characteristics.

decade. The InGaAs nanowire SGT had an on/off current ratio (I_{ON}/I_{OFF}) of 10^6 and a peak transconductance (G_m) of 280 μS μm⁻¹ at a drain–source voltage of $V_{DS} = 1.00$ V. The drain-induced barrier lowering (DIBL) was estimated to be 48 mV V⁻¹. After analysing InGaAs nanowires SGTs with various EOTs (Supplementary Fig. 6), we found that SS and the DIBL were independent of the EOT and averaged 82 mV dec⁻¹ and 45 mV V⁻¹, respectively. These switching characteristics are much better than those of SGTs using InAs nanowire channels on Si (refs 21–24); however, this simple, InGaAs nanowire-based SGT requires further improvements to satisfy the technological demands of next-generation metal–oxide–semiconductor FETs². In particular, the channel mobility must be enhanced by decreasing the number of surface states.

We therefore designed and fabricated an improved structure, that is, In_{0.7}Ga_{0.3}As/InP/In_{0.5}Al_{0.5}As/ δ -doped In_{0.5}Al_{0.5}As/In_{0.5}Al_{0.5}As/In_{0.7}Ga_{0.3}As modulation-doped core–multishell (MD-CMS) nanowires (Fig. 3a). The advantage of selective-area growth is the ability to form a CMS structure that allows surface passivation¹⁹. In addition, modulation-doped layers such as those on high-electron-mobility transistors²⁵ can be packed into the CMS layers. The InGaAs outer-shell layer is a cap designed to maintain good interface quality with the Hf_{0.8}Al_{0.2}O layer, and confines carriers when $V_G < V_T$. The InAlAs/ δ -doped layer/InAlAs structure forms the modulation-doped layers. The I_D and V_T depends on the confined carriers and the region to which they are confined. The InP layer is a barrier that allows the position of the

confined carriers to be controlled by varying V_G . These confined electrons do not contribute to the drain current because this layer is separated from the drain metal. However, electrons penetrate the InP barrier layer and are confined in the core InGaAs nanowire when $V_G > V_T$. The band diagram of the MD-CMS nanowire when $V_G = 0.50$ V (Fig. 3b, calculated using the one-dimensional Poisson–Schrödinger equation²⁶) suggests that the carriers are confined in the nanowire for, in that case, $V_G = 0.50$ V. The MD-CMS nanowire SGT is thus normally off. Figure 3c is a representative SEM image of the vertical MD-CMS nanowires on Si. The HAADF-STEM image and EDX elemental mappings in Figs 3d–j show that the CMS layers

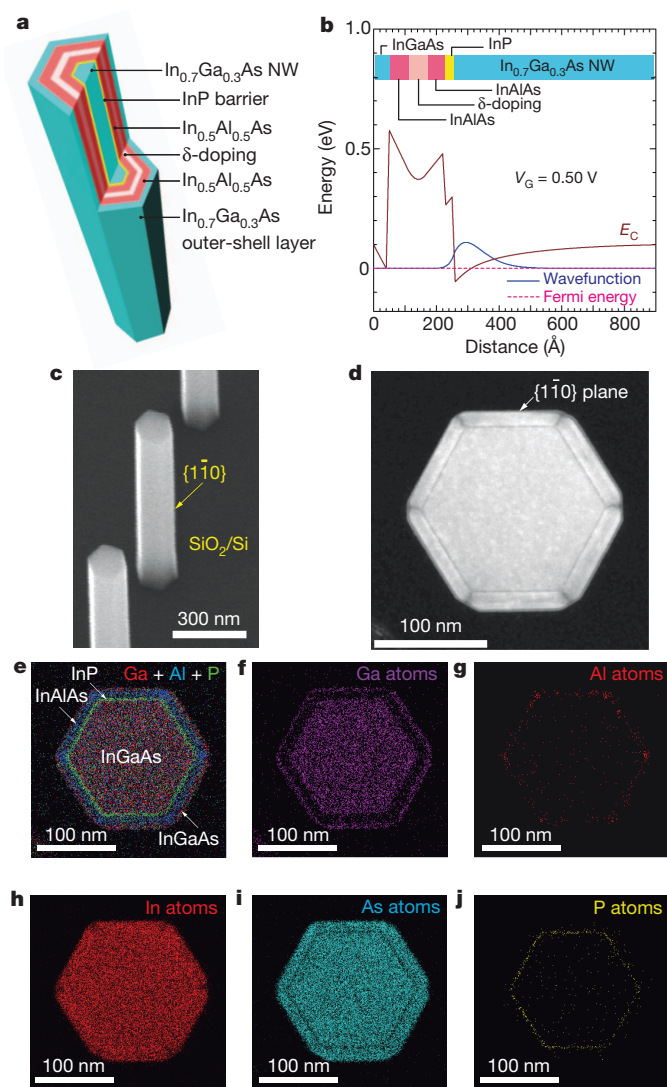


Figure 3 | Formation and characterization of InGaAs/InP/InAlAs/InGaAs CMS nanowires on Si. **a**, Illustration of designed and fabricated structure. The InGaAs nanowire core is wrapped with InP/InAlAs/ δ -doped InAlAs/InAlAs/InGaAs multilayers. **b**, Band diagram of the MD-CMS nanowire at $V_G = 0.50$ V simulated using the one-dimensional Poisson–Schrödinger equation, suggesting confinement of carrier wavefunction (blue curve) in the InGaAs nanowire. The red dashed line denotes the Fermi level. The conduction band is denoted E_C . Without a gate bias, the carrier wavefunction is confined in the outer InGaAs shell layer. **c**, Typical SEM image showing vertical MD-CMS nanowires on Si. The average diameter and height of these nanowires are 180 nm and 1.2 μm , respectively. The CMS layers are grown on the side walls of the core InGaAs nanowires. **d**, HAADF-STEM image showing a representative cross-section of a CMS nanowire with a total diameter of 180 nm. **e–j**, EDX elemental mapping images: mixing of Ga, Al and P (**e**); Ga (**f**); Al (**g**); In (**h**); As (**i**); and P (**j**). The Al mapping (**g**) shows segregation of the Al at the corner of the CMS nanowire.

formed around the InGaAs nanowire core. The thicknesses of the InP, InAlAs, δ -doped, InAlAs and InGaAs capping layers were respectively about 2.6, 5.5, 5.5, 5.5 and 5.0 nm (Supplementary Fig. 7). The dark regions at the corners of the outer MD-CMS layer (Fig. 3d) indicate that Al-rich InAlAs formed through segregation of Al. The Al-rich parts separate the tubular modulation-doped channels into six-sided transistor layers. Accordingly, the six-sided transistor layers could be integrated onto a single tiny nanowire.

Figure 4a, b shows the device performance of the SGTs using MD-CMS nanowire channels on Si. The diameter of the InGaAs nanowire cores was 90 nm. The drain currents shown are those for a single wire (measured current divided by the number of nanowires and normalized by gate outer perimeter). The SGT showed an n-type enhancement mode with a threshold voltage of 0.38 V. The $C-V_G$ characteristic (Fig. 4c) shows a positive shift of the flat-band voltage and a reduction in capacitance due to the gate dielectric/CMS layers, relative to that of the InGaAs nanowire SGT. The $I_{\text{ON}}/I_{\text{OFF}}$ ratio was approximately 10^8 . The I_{OFF} was $<10 \text{ pA } \mu\text{m}^{-1}$, which was much lower than that of a conventional Si metal–oxide–semiconductor FET¹. The SS and DIBL were 75 mV per decade and $35 \text{ mV } \text{V}^{-1}$, respectively. The CMS nanowire SGTs achieved both a steep SS and a very low DIBL, indicating suppression of the short-channel effect owing to the multigate architecture¹. The drain current reached $0.45 \text{ mA } \mu\text{m}^{-1}$ at $V_G - V_T = 0.50$ V, and G_m was $1.42 \text{ mS } \mu\text{m}^{-1}$ at $V_{\text{DS}} = 0.50$ V. The very large increase in I_D and G_m together with the steep SS indicates that the CMS multilayers acted as a modulation-doped, high-electron-mobility structure while maintaining good controllability of the surrounding-gate structure. These results are superior to those from

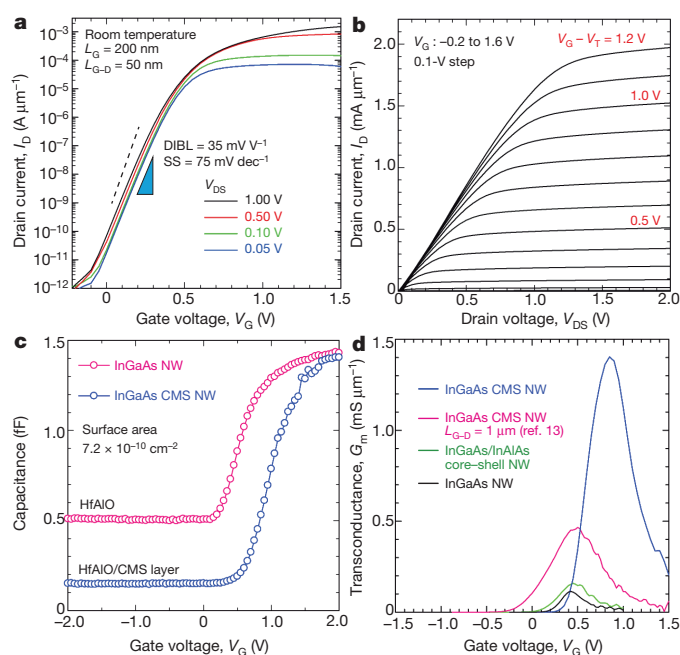


Figure 4 | Performance of an SGT using InGaAs/InP/InAlAs/InGaAs CMS nanowire channels on Si. **a**, Transfer characteristics of an SGT with $L_G = 200$ nm and $L_{G-D} = 50$ nm. SS is 75 mV per decade and DIBL is $35 \text{ mV } \text{V}^{-1}$. The dashed line indicates the physical limit of SS (60 mV per decade at room temperature). **b**, Output characteristics of an InGaAs CMS nanowire SGT. **c**, Capacitance/gate voltage curve of a InGaAs nanowire (pink circles) and a InGaAs CMS nanowire (blue circles). The capacitances shown are the measured values divided by the number of nanowires (250 for InGaAs nanowires and 512 for InGaAs CMS nanowires). The frequency in this measurement is 1 MHz. **d**, Transconductance of SGTs using InGaAs nanowires (black), InGaAs/InAlAs core-shell nanowires (green), MD-CMS nanowires with $L_{G-D} = 1 \text{ } \mu\text{m}$ (pink; ref. 13) and MD-CMS nanowires (blue). $V_{\text{DS}} = 0.50$ V.

devices made from similar materials and with similar dimensions^{12,13}. Figure 4d shows G_m curves for the SGTs using InGaAs nanowires, InGaAs/InAlAs core-shell nanowires and MD-CMS nanowires. The MD-CMS nanowire channel results in a pronounced increase in G_m . The field effect mobility tentatively estimated from G_m (Supplementary Fig. 8) is much higher than the typical electron mobility of a Si metal-oxide-semiconductor FET²⁷. Consequently, the SGT devices reported here could have the performance necessary for use in future Si transistor technology²⁸.

METHODS SUMMARY

The InGaAs nanowires and InGaAs/InP/InAlAs/InGaAs CMS nanowires were grown by using selective-area metal-organic vapour phase epitaxy. High-resolution TEM (H-9000UHR with acceleration voltage of 300 kV; Hitachi) was used to evaluate the crystal structure and to estimate the strain²⁹. HAADF-STEM (JEM2100F with an acceleration voltage of 200 kV) and EDX (JEOL JED-2300T) were used to analyse the elemental mapping of the CMS nanowires. The SGTs were fabricated with a reactive-ion etching procedure. The current-voltage curves were measured with an Agilent 4192A impedance analyser. The transistor performance was evaluated with an Agilent 4156C parameter analyser. All current measurements were performed at room temperature (300 K) in the dark. For display, the measured capacitances were divided by the number of nanowires and the measured currents were divided by the number of nanowires and normalized by the gate outer perimeter.

Full Methods and any associated references are available in the online version of the paper.

Received 9 March; accepted 31 May 2012.

Published online 1 August 2012.

1. Ferain, I., Colinge, A. A. & Colinge, J.-P. Multigate transistors as the future of classical metal-oxide-semiconductor field-effect transistors. *Nature* **479**, 310–316 (2011).
2. del Alamo, J. A. Nanometre-scale electronics with III–V compound semiconductors. *Nature* **479**, 317–323 (2011).
3. Seabaugh, A. C. & Zhang, Q. Low-voltage tunnel transistors for beyond CMOS logic. *Proc. IEEE* **98**, 2095–2110 (2010).
4. Ionescu, A. M. & Riel, H. Tunnel field-effect transistors as energy-efficient electronic switches. *Nature* **479**, 329–337 (2011).
5. Radosavljevic, M. *et al.* Advanced high- k gate dielectric for high-performance short-channel $\text{In}_{0.7}\text{Ga}_{0.3}\text{As}$ quantum well field effect transistors on silicon substrate for low power logic applications. *IEDM Tech. Dig.* 319–322 (2009).
6. Kim, S. H. *et al.* Electron mobility enhancement of extremely thin body $\text{In}_{0.7}\text{Ga}_{0.3}\text{As}$ -on-insulator metal-oxide-semiconductor field-effect transistors on Si substrates by metal-oxide-semiconductor interface buffer layer. *Appl. Phys. Exp.* **5**, 014201 (2012).
7. Wu, Y. Q. *et al.* 0.8-V supply voltage deep-submicrometer inversion-mode $\text{In}_{0.75}\text{Ga}_{0.25}\text{As}$ MOSFET. *IEEE Elec. Dev. Lett.* **30**, 700–702 (2009).
8. Radosavljevic, M. *et al.* Non-planar, multi-gate InGaAs quantum well field effect transistors with high- k gate dielectric and ultra-scaled gate-to-drain/gate-to-source separation for low power logic applications. *IEDM Tech. Dig.* 126–129 (2010).
9. Radosavljevic, M. *et al.* Electrostatic improvement in 3-D tri-gate over ultra-thin body planar InGaAs quantum well field effect transistors with high- k gate dielectric and scaled gate-to-drain/gate-to-source separation. *IEDM Tech. Dig.* 765–768 (2011).
10. Wieder, H. H. Surface and interface barriers of $\text{In}_x\text{Ga}_{1-x}\text{As}$ binary and ternary alloys. *J. Vac. Sci. Technol. B* **21**, 1915–1919 (2003).
11. Takato, H. *et al.* Impact of surrounding gate transistor (SGT) for ultra-high-density LSI's. *IEEE Trans. Electron. Dev.* **38**, 573–578 (1991).
12. Gu, J. J. *et al.* First experimental demonstration of gate-all-around III–V MOSFETs by top-down approach. *IEDM Tech. Dig.* 769–772 (2011).
13. Tomioka, K., Yoshimura, M. & Fukui, T. Vertical $\text{In}_{0.7}\text{Ga}_{0.3}\text{As}$ nanowire surrounding-gate transistors with high- k gate dielectrics on Si substrate. *IEDM Tech. Dig.* 773–776 (2011).
14. Xuan, Y. *et al.* Submicrometer inversion-type enhancement-mode InGaAs MOSFET with atomic-layer-deposited Al_2O_3 as gate dielectric. *IEEE Elec. Dev. Lett.* **28**, 935–938 (2007).
15. Ishii, H. *et al.* High electron mobility metal-insulator-semiconductor field-effect transistors fabricated on (111)-oriented InGaAs channels. *Appl. Phys. Exp.* **2**, 121101 (2009).
16. Kim, D.-H. *et al.* 50-nm E-mode $\text{In}_{0.7}\text{Ga}_{0.3}\text{As}$ PHEMTs on 100-nm InP substrate with $f_{\text{max}} > 1$ THz. *IEDM Tech. Dig.* 692–695 (2010).
17. Furukawa, Y. *et al.* Control of N content of GaPN grown by molecular beam epitaxy and growth of GaPN lattice matched on Si(100) substrate. *Jpn. J. Appl. Phys.* **41**, 528–532 (2002).
18. Tomioka, K., Motohisa, J., Hara, S. & Fukui, T. Control of InAs nanowire growth directions on Si. *Nano Lett.* **8**, 3475–3480 (2008).
19. Tomioka, K. *et al.* Selective-area growth of vertically aligned GaAs and GaAs/AlGaAs core-shell nanowires on Si(111) substrate. *Nanotechnology* **20**, 145302 (2009).
20. Noborisaka, J. *et al.* Electrical characterization of InGaAs nanowire-top-gate field-effect transistors by selective-area metal organic vapor phase epitaxy. *Jpn. J. Appl. Phys.* **46**, 7562–7568 (2007).
21. Rehnsdt, C. *et al.* Vertical InAs nanowire wrap gate transistors on Si substrate. *IEEE Trans. Electron. Dev.* **55**, 3037–3041 (2008).
22. Tanaka, T. *et al.* Vertical surrounding gate transistors using single InAs nanowires grown on Si substrates. *Appl. Phys. Exp.* **3**, 025003 (2010).
23. Wernersson, L.-E., Thelander, C., Lind, E. & Samuelson, L. III–V nanowires—extending a narrowing road. *Proc. IEEE* **98**, 2047–2060 (2010).
24. Ghalamestani, S. G. *et al.* Uniform and position controlled InAs nanowires on 2'' Si substrate for transistor applications. *Nanotechnology* **23**, 015302 (2012).
25. Mimura, T., Hiyamizu, S., Fujii, T. & Nanbu, K. A new field-effect transistor with selectively doped GaAs/n-Al_xGa_{1-x}As heterojunctions. *Jpn. J. Appl. Phys.* **19**, L225–L227 (1980).
26. Tan, I.-H., Snider, G. L., Chang, L. D. & Hu, E. L. A self-consistent solution of Schrodinger-Poisson equations using nonuniform mesh. *J. Appl. Phys.* **68**, 4071–4076 (1990).
27. Takagi, S., Toriumi, A., Iwase, M. & Tango, H. On the universality of inversion layer mobility in Si MOSFET's: part I—effects of substrate impurity concentration. *IEEE Trans. Electron. Dev.* **41**, 2357–2362 (1994).
28. ITRS. *International Technology Roadmap for Semiconductors* <http://www.itrs.net/Links/2011ITRS/Home2011.htm> (ITRS, 2011).
29. Galindo, P. L. *et al.* The peak pairs algorithm for strain mapping from HRTEM images. *Ultramicroscopy* **107**, 1186–1193 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Motohisa, S. Hara, K. Hiruma, T. Hashizume and T. Waho for discussions. We especially thank Y. Hori for instructing us on the current-voltage measurement. This work was financially supported by a Grant-in-Aid for Scientific Research from MEXT and the Japan Science and Technology Agency – PRESTO programme.

Author Contributions K.T. designed the experiments, made the nanowires by metal-organic vapour phase epitaxy, fabricated the device and analysed all of the data. T.F. planned and supervised the study. M.Y. helped in the epitaxy experiments. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.T. (tomioka@rciqe.hokudai.ac.jp).

METHODS

Selective-area metal–organic vapour phase epitaxy. After the Si(111) substrate was degreased with organic solvents, a 20-nm-thick SiO₂ film was formed by thermal oxidation. Circular openings arranged in a triangular lattice with a pitch of 3 µm were then formed on the SiO₂ films by using electron-beam lithography and wet chemical etching. The opening diameter, d_0 , was 90 nm. The nanowires were grown using metal–organic vapour phase epitaxy.

Growth of InGaAs nanowires. The InGaAs nanowires were grown in a horizontal, low-pressure (0.1-atm) system. Trimethylgallium (TMGa), trimethylindium (TMIn) and arsine (AsH₃) gas were used as material sources and monosilane (SiH₄) was used as the n-type dopant. After thermal cleaning at 900 °C in H₂, AsH₃ treatment at 670 °C and flow-rate modulation epitaxy were carried out, in which the material source (TMGa + TMIn) and AsH₃ were supplied alternately with intervals of H₂. The TMGa + TMIn, AsH₃ and H₂ supply durations were respectively 1, 1 and 2 s. After that, InGaAs nanowires were grown for 20 min at 670 °C. The partial pressures of the TMGa, TMIn, AsH₃ and SiH₄ were respectively 5.7×10^{-7} , 9.7×10^{-7} , 2.5×10^{-4} and 2.5×10^{-7} atm.

Formation of InGaAs/InP/InAlAs/InGaAs CMS nanowires. After the growth of the InGaAs nanowires, InP/InAlAs/ δ -doped InAlAs/InAlAs/InGaAs layers were grown at 580 °C. Trimethylaluminium (TMAI) and *tert*-butylphosphine (TBP) were used as the Al and P sources. The growth times for the InP, InAlAs, δ -doped InAlAs, InAlAs and InGaAs layer were respectively 10, 40, 40, 40 and 30 s. The partial pressures of the TMIn and TBP for the InP layer were respectively 4.4×10^{-6} and 1.6×10^{-4} atm. The partial pressures of the TMAI, TMIn, AsH₃ and SiH₄ for the InAlAs layer were respectively 5.3×10^{-7} , 4.9×10^{-7} , 2.5×10^{-4} , 2.5×10^{-4} and 1.2×10^{-7} atm.

Strain mapping. Strain mapping estimated from the displacement of bright spots in the TEM image is shown in Fig. 1f. The strains, ϵ_{xx} and ϵ_{yy} (Supplementary Fig. 3), were calculated from the displacement of bright spots in Fig. 1d by using a peak-pair

algorithm²⁹, and the displacements of the bright spots are defined by $u_{xx} = \Delta x - a_{\text{Si}(x)}$ for the in-plane $\langle 2\bar{1}1 \rangle$ direction and by $u_{yy} = \Delta y - a_{\text{Si}(y)}$ for the vertical $\langle 111 \rangle$ direction. Here Δx and Δy are the displacements of the bright spots in each direction, and $a_{\text{Si}(x)}$ and $a_{\text{Si}(y)}$ are the lattice constants in the in-plane and vertical directions of the Si(111) substrate, estimated from the TEM image. The strains ϵ_{xx} and ϵ_{yy} are also given by $\epsilon_{xx} = \partial u / \partial x$ and $\epsilon_{yy} = \partial u / \partial y$, where $u = \sqrt{u_{xx}^2 + u_{yy}^2}$. We note that, because the displacement of the atoms is calculated on the basis of the position of the atoms in crystalline Si, InGaAs is mapped into a layer with a strain of +8.1% by definition. The error in the strain calculation is approximately $\pm 0.5\%$.

Fabrication of SGTs. After the InGaAs nanowire growth, the nanowires were treated with an alkaline solution to etch away native oxides. The nanowires were then covered with Hf_{0.8}Al_{0.4}O_x ($\epsilon_{\text{HfAlO}} = 20.4$) film using atomic layer deposition. This oxide was used as the gate oxide, and it ranged in thickness from 10 nm (EOT = 1.86 nm) to 20 nm (EOT = 3.72 nm). Next the gate metal, tungsten (W), was deposited by radio-frequency sputtering (Supplementary Fig. 4b). After the W was lithography patterned for nanowire-grown masks ($50 \times 50 \mu\text{m}^2$), the nanowires were spin-coated with benzocyclobutene (BCB) (Supplementary Fig. 4c) and etched back by reactive-ion etching (RIE) with CF₄/O₂ to etch the BCB, W and Hf_{0.8}Al_{0.2}O gate oxide simultaneously (Supplementary Fig. 4d). After the RIE process, the nanowires were spin-coated with BCB, and etched back again by RIE (Supplementary Fig. 4e) to isolate the gate and drain metals. A Ni/Ge/Au/Ni/Au multilayer was evaporated onto a lithographically defined region to serve as the drain contact. A Ti/Au multilayer was deposited onto the Si substrate to serve as the source contact (Supplementary Fig. 4f). The device had ten nanowires, which were connected in parallel to a single drain contact pad. The gate length was 200 nm. Finally, the nanowire SGT was annealed at 420 °C in N₂ to obtain ohmic contacts at the source and drain regions.

A new atmospherically relevant oxidant of sulphur dioxide

R. L. Mauldin III^{1,2,3}, T. Berndt⁴, M. Sipilä^{1,4,5}, P. Paasonen¹, T. Petäjä¹, S. Kim², T. Kurtén^{1,6}, F. Stratmann⁴, V.-M. Kerminen¹ & M. Kulmala¹

Atmospheric oxidation is a key phenomenon that connects atmospheric chemistry with globally challenging environmental issues, such as climate change¹, stratospheric ozone loss², acidification of soils and water³, and health effects of air quality⁴. Ozone, the hydroxyl radical and the nitrate radical are generally considered to be the dominant oxidants that initiate the removal of trace gases, including pollutants, from the atmosphere. Here we present atmospheric observations from a boreal forest region in Finland, supported by laboratory experiments and theoretical considerations, that allow us to identify another compound, probably a stabilized Criegee intermediate (a carbonyl oxide with two free-radical sites) or its derivative, which has a significant capacity to oxidize sulphur dioxide and potentially other trace gases. This compound probably enhances the reactivity of the atmosphere, particularly with regard to the production of sulphuric acid, and consequently atmospheric aerosol formation. Our findings suggest that this new atmospherically relevant oxidation route is important relative to oxidation by the hydroxyl radical, at least at moderate concentrations of that radical. We also find that the oxidation chemistry of this compound seems to be tightly linked to the presence of alkenes of biogenic origin.

Oxidation of trace gases drives atmospheric chemistry and influences thereby both air quality and climate, and their interaction with each other and the biosphere^{1,5–7}. The main gas-phase oxidants under consideration (so far) are the OH radical (OH•, referred to here as OH for simplicity), O₃, the nitrate radical (NO₃•, referred to here as NO₃ for simplicity) and Cl atoms, of which OH is important only during the daytime and NO₃ during night time. Recently, an increasing number of investigations have focused on atmospheric reactivity, more specifically the missing reactivity^{8,9} and sources^{5,10} of OH and its temporal variability¹¹, as well as missing HONO sources⁷. Sulphuric acid (H₂SO₄) is a key compound, connecting atmospheric oxidation chemistry with the formation and growth of new aerosol particles¹². Until now, the general consensus has been that the rate at which sulphur dioxide (SO₂) is converted to gaseous H₂SO₄ is determined by the OH concentration. Here we show that there is another important source of gaseous H₂SO₄ that is not directly related to OH.

Our investigation into this uncharted oxidation chemistry is based on simultaneous observations of OH and H₂SO₄ using chemical ionization mass spectrometry¹³ (CIMS; see Supplementary Information). When measuring OH concentrations using this technique, OH is first converted to isotopically labelled H₂SO₄ by the addition of ³⁴SO₂ to the ambient sample flow; the resulting H₂³⁴SO₄ is then measured using CIMS. There are, however, other processes which can oxidize SO₂ to H₂SO₄. To determine which fraction of the measured H₂³⁴SO₄ originates from the reaction of atmospheric OH with ³⁴SO₂, the measurement is repeated with the addition of an OH scavenger (propane) to the sample flow to suppress H₂³⁴SO₄ formation via OH. This procedure results in a 'background' H₂³⁴SO₄

concentration (produced from X + SO₂), termed [X] here, which is subtracted from the total H₂³⁴SO₄ concentration to obtain the ambient OH concentration. We hypothesize that [X] represents a yet unexplored oxidant X (or sum of several such oxidants) which, similarly to OH, is capable of converting SO₂ to gaseous sulphuric acid in the atmosphere.

We started our investigation using field observations performed at the SMEAR II station in the Finnish boreal forest region (Supplementary Information). Figure 1a and b shows the concentration time series of [OH], [X] and [H₂SO₄] measured over one week during the summer of 2010. The OH concentration shows a typical diurnal cycle¹⁴, with maximum concentrations around noon and much lower ones during the night. The value of [X] does not show a clear diurnal cycle, but it typically exceeds [OH]. During several evenings and nights, we identify instances when [OH] is close to 10⁵ molecules cm⁻³, [X] simultaneously exceeds 10⁶ molecules cm⁻³, and [H₂SO₄] is remarkably high, up to about 10⁶ molecules cm⁻³. This observation indicates the presence of a non-OH source for H₂SO₄ production, and further suggests that there might be a connection between this source and the oxidant X. We calculated the H₂SO₄ concentration resulting from the reaction of SO₂ with OH (green line in Fig. 1b). The difference between the H₂SO₄ concentration measured by the CIMS and that due to the reaction of SO₂ with OH is our best estimate of the H₂SO₄ concentration resulting from the non-OH source, [H₂SO₄]_{non-OH}. Figure 1c shows that the value of [H₂SO₄]_{non-OH} increases with increasing [X], reaching values as high as (2–3) × 10⁶ molecules cm⁻³ during our measurements. Sulphuric acid originating from this non-OH source may contribute up to 50% of the total H₂SO₄ budget (Fig. 1b and Supplementary Information), demonstrating the important role of this H₂SO₄ formation route.

The dominance of [X] over [OH] particularly during evenings and nights suggests that the compound X might be related to surface emissions and subsequent ozone chemistry taking place in the boundary layer. In order to investigate this chemistry, we carried out laboratory experiments using two systems with different flow characteristics (Supplementary Information). In these experiments, SO₂ was exposed to mixtures of ozone and various alkenes, and the resulting H₂SO₄ concentration was measured with CIMS and modelled using a scheme based on known OH chemistry (Supplementary Information). Because alkene–ozone reactions are known to produce OH, the experiments were conducted both with and without an OH-scavenger (H₂ or CO). Figure 2a shows the measured and modelled H₂SO₄ concentration as a function of the amount of monoterpene (limonene and α-pinene) reacted with ozone. As in the field study, we observed H₂SO₄ that cannot be explained by the reaction of SO₂ with OH alone. In these experiments, the production of H₂SO₄ from this non-OH source appeared to be more efficient for monoterpenes than for other alkenes (for example, MCH, 1-methyl-cyclohexene; see Supplementary Information). The role of the new H₂SO₄ production chemistry

¹University of Helsinki, Department of Physics, FI-00014 Helsinki, Finland. ²National Center for Atmospheric Research, Boulder, Colorado 80307, USA. ³Department of Atmospheric and Oceanic Sciences, University of Colorado at Boulder, Boulder, Colorado 80309, USA. ⁴Leibniz-Institute for Tropospheric Research, 04318 Leipzig, Germany. ⁵Helsinki Institute of Physics, FI-00014 Helsinki, Finland.

⁶University of Helsinki, Department of Chemistry, FI-00014 Helsinki, Finland.

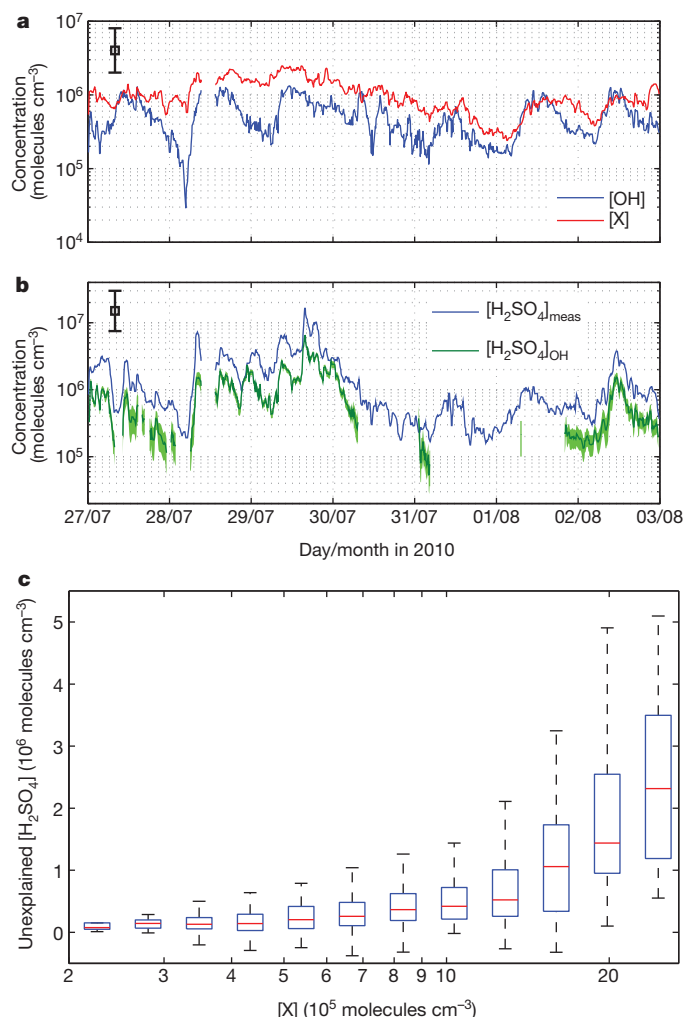


Figure 1 | Data from a boreal forest site, obtained during summer 2010. **a**, Time series of $[\text{OH}]$ (blue line) and of $[\text{X}]$ (red line); **b**, time series of the H_2SO_4 concentration measured by CIMS (blue line), and of the H_2SO_4 concentration attributable to the reaction of OH with SO_2 , $[\text{H}_2\text{SO}_4]_{\text{OH}}$ (green line); data were obtained from 27 July to 3 August 2010 at Hyytiälä, Finland. The value of $[\text{H}_2\text{SO}_4]_{\text{OH}}$ is determined on the basis of the measured SO_2 and OH concentrations and the calculated condensation sink for gaseous H_2SO_4 , ref. 30. **c**, Estimated concentration of H_2SO_4 attributable to the non- OH source as a function of $[\text{X}]$. Red lines are median values; boxes depict 25th and 75th percentiles; black bars enclose all data excluding outliers. We now consider uncertainties in **a** and **b**. The vertical bars in the upper left corners of **a** and **b** illustrate the uncertainty range (\pm a factor of 2) in the measured values of $[\text{H}_2\text{SO}_4]$ and $[\text{OH}]$, as well as in measured $[\text{X}]$ under the assumption that X is fully converted to $[\text{H}_2^{34}\text{SO}_4]$. These are absolute uncertainties, based on the stated uncertainties of the values used in the calculation (such as CIMS calibration lamp intensity, water photolysis reaction rate coefficient and OH losses). The statistical uncertainties are less important in comparison with these uncertainties. The shaded green area in **b** depicts the range of additional uncertainty in calculated $[\text{H}_2\text{SO}_4]_{\text{OH}}$ obtained by taking into account the uncertainties in the measured values of the SO_2 concentration (± 0.05 p.p.b.) and condensation sink. The upper limit of each error estimate was calculated by assuming that the SO_2 concentration was 0.05 p.p.b. larger than the measured value and that the condensation sink was at its minimum value at the measured relative humidity. The minimum value of the condensation sink, CS , was obtained by assuming that there were no super-micrometre particles and that the hygroscopicity of the sub-micrometre particles was at its minimum. Consequently, the lower limit of each error estimate was calculated by assuming that the SO_2 concentration was 0.05 p.p.b. lower than the measured value, and that the condensation sink was at its maximum value at the measured relative humidity. The maximum value of CS was obtained by assuming that the super-micrometre particles contributed 5% to total CS , and that the hygroscopicity of the sub-micrometre particles was at its maximum (see Supplementary Information and references therein).

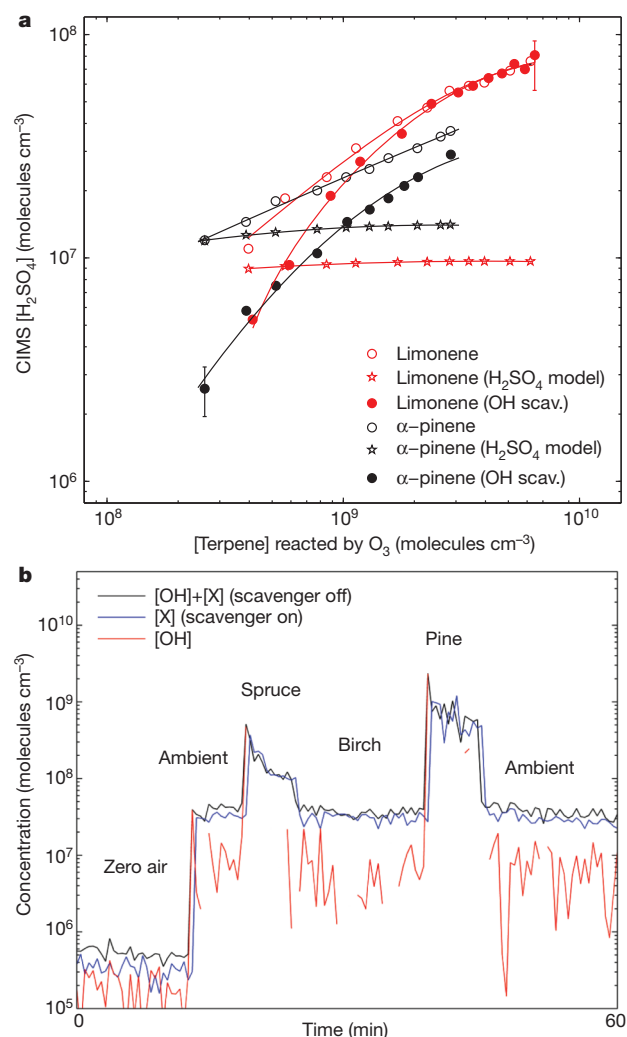


Figure 2 | Plots showing that alkenes emitted by vegetation are possibly the source of oxidant X. Measured H_2SO_4 concentrations as a function of reacted terpene concentration (α -pinene or limonene) in the presence and absence of an OH scavenger (H_2) as a result of IFT-LFT laboratory experiments. Data were obtained using constant O_3 and SO_2 concentrations (O_3 , 24.7 p.p.b.v.; SO_2 , 14.6 p.p.b.v.), and $[\text{H}_2]$ was 2.4×10^{17} molecules cm^{-3} . Modelling results (including only H_2SO_4 formation via $\text{OH} + \text{SO}_2$) are shown for runs in the absence of the OH scavenger; compare the reaction scheme shown in Supplementary Information. The results from limonene when compared with those from α -pinene clearly show that the former is more efficient at producing the additional H_2SO_4 . Error bars are $\pm 35\%$ and indicate the uncertainty in H_2SO_4 measurements; they represent an absolute uncertainty resulting from a propagation of uncertainties in the CIMS calibration lamp intensity, water photolysis reaction rate coefficient and OH losses (see details in Supplementary information). **b**, Pine and spruce are potentially major species emitting precursors of X . Here data series show the measured concentration of all oxidants (OH and X) capable of oxidizing SO_2 in absence of an OH scavenger, the measured concentration of X in the presence of an OH scavenger, as well as the difference between the two, which corresponds to the concentration of OH . Data were obtained in Hyytiälä, a boreal forest site in late summer (5 September 2011). Zero air measurements indicate the level of instrumental background. Measurements were performed by placing the branches from different tree varieties in the direct vicinity of the instrument's inlet, one species at the time. VOCs emitted by the branches react with ambient O_3 , subsequently producing OH and sCI . As $[\text{OH}]$ is calculated from the difference of measured $[\text{OH}] + [\text{X}]$ and $[\text{X}]$, $[\text{OH}]$ cannot be accurately determined during high $[\text{X}]$ (pine and spruce). Both spruce and pine yielded enormous increases in signal, while birch had no observable effect (see Supplementary Information for a complete figure). Cutting of branches results in enlarged emissions of monoterpenes from trees with large storage pools, such as spruce (*Picea abies*) and pine (*Pinus sylvestris*)¹⁵.

becomes dominant at high monoterpene concentrations, as shown by the convergence of the data series taken in the absence and in the presence of the OH scavenger.

Monoterpenes, including limonene and α -pinene used in our experiments, are emitted effectively by trees, and these compounds are abundantly present at our field measurement site during the summertime^{15,16}. To confirm vegetation as a source of the alkenes responsible for X formation in the boreal forest environment, we performed an additional experiment where branches of different trees were cut and placed in the immediate vicinity of the CIMS inlet (Fig. 2b, see also Supplementary Information). The production of OH from ozonolysis of branch emissions during this experiment was minor in comparison to production of X. This experiment indisputably substantiates our conclusion, demonstrating the role of trees in producing compound X and, consequently, affecting gaseous sulphuric acid production.

The field and laboratory measurements presented above give strong evidence of the existence of a previously unknown oxidant X, but do not reveal its identity. Experiments^{17,18} and quantum chemical calculations^{19,20} have demonstrated that the reactions of SO₂ with the most common non-OH oxidants (O₃ and NO₃) and with peroxy radicals (HO₂, H₃COO and larger analogues) are extremely slow. Stabilized Criegee intermediates (sCIs), formed in the ozonolysis of all alkenes, are known to oxidize SO₂ (ref. 21) but the rate constant of the SO₂ + sCI reaction has been assumed to be fairly low, of the order of 10^{-15} cm³ s⁻¹ (ref. 22). However, recent theoretical quantum mechanical studies^{20,23}, as well as laboratory experiments²⁴, have found the SO₂ + sCI reaction to be significantly faster than previously thought. Another property that influences the oxidation capacity of sCIs is their lifetime against unimolecular decomposition reactions. The sCIs formed in ozonolysis of larger alkenes, such as monoterpenes, may have longer lifetimes than those formed from lower-molecular-weight alkenes. We note here that besides sCIs, other ozonolysis intermediates might also be responsible for the observed additional oxidation of SO₂ (refs 25, 26). Figure 3 summarizes schematically this new mechanism of atmospheric oxidation chemistry.

We finalize our analysis by investigating whether the proposed chemistry is consistent with field observations. For this purpose, we estimated the reaction rate between sCI and SO₂ from the laboratory measurement data presented in Fig. 2a (see Supplementary Information for theoretical considerations and assumptions made in deriving the reaction rate). The resulting rate coefficient was about 6×10^{-13}

cm³ s⁻¹ for α -pinene + sCI and about 8×10^{-13} cm³ s⁻¹ for limonene + sCI. In Fig. 4a and b we have calculated the H₂SO₄ concentration originating from the two SO₂ oxidation pathways, OH and non-OH, during the field measurement period reported in Fig. 1. The comparison of the calculated values of [H₂SO₄] with the measured ones shows that in addition to explaining the large missing H₂SO₄ source

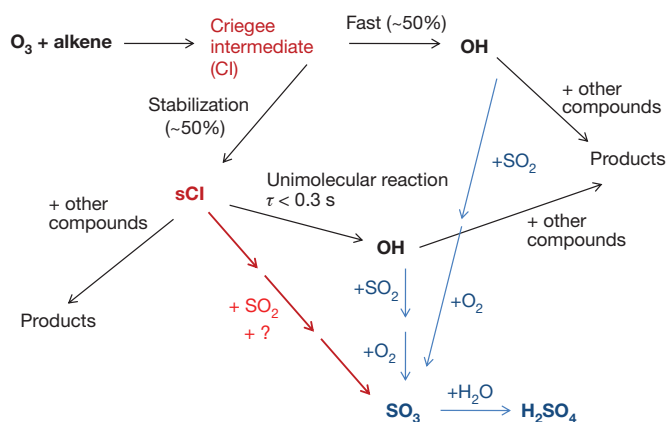


Figure 3 | Proposed mechanism for the formation of oxidant X. Of the Criegee intermediates (CI) formed during ozonolysis, ~50% decompose to produce OH on a subsecond timescale while the other 50% are stabilized, producing stabilized Criegee radicals, sCI. These sCI can then decompose over a much longer lifetime, τ . Our observations suggest that sCI, or non-OH derivatives of sCI, can also oxidize at least SO₂ (red arrows), thus altering the known view of oxidation chemistry in the atmosphere. Typical OH chemistry is depicted in blue.

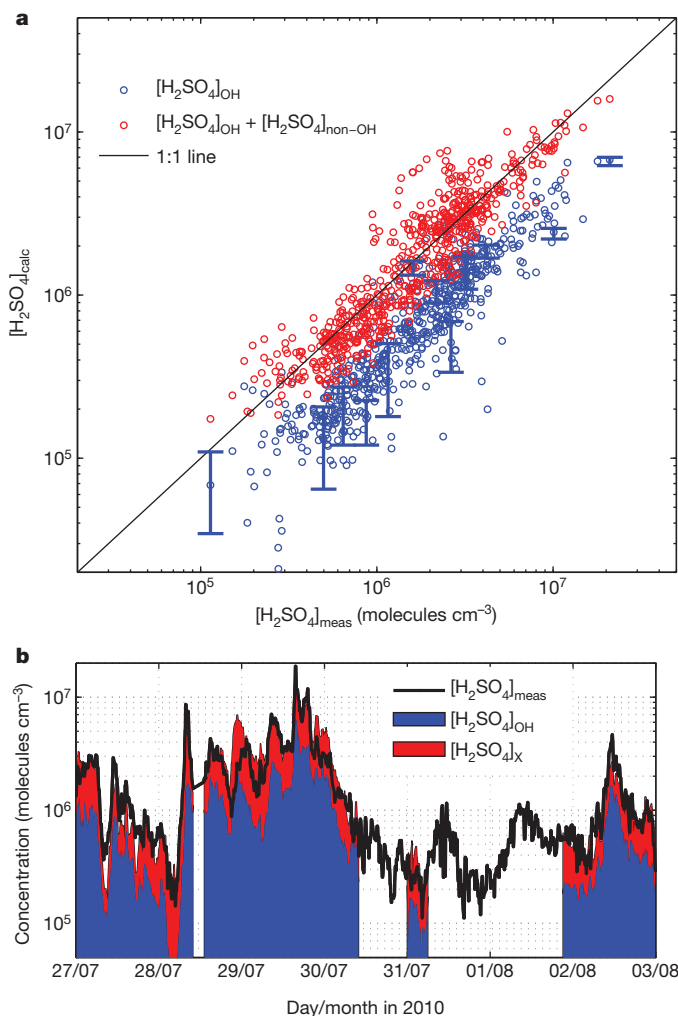


Figure 4 | H₂SO₄ produced from X + SO₂ can explain the difference between the Hyytiälä 2010 H₂SO₄ measurements and calculated values using only OH + SO₂. **a**, Calculated sulphuric acid concentration as a function of the measured concentration. Blue circles represent the concentration calculated for SO₂ oxidation solely via OH, whereas the red circles correspond to the concentration calculated for SO₂ oxidation by both OH and X. The reaction rate coefficient for OH oxidation was taken from literature³⁰ and that for X oxidation determined from the laboratory experiments (Fig. 2) described in the text and Supplementary Information. The error in [H₂SO₄]_{OH} is estimated as described for the shaded green area shown in Fig. 1, and for clarity is plotted for a few data points only. This error cannot explain the observed systematic difference between [H₂SO₄]_{OH} and [H₂SO₄]_{meas} (the average ratio between these quantities being 0.41). The error in [H₂SO₄]_{non-OH} cannot be accurately quantified and therefore no error estimate for the red data points has been given. The errors resulting from uncertainties in measured OH or H₂SO₄ concentrations do not affect the position of blue or red points with respect to the one-to-one line. Correlation coefficients (log-log scale) for the two data series are $R^2 = 0.81$ and $R^2 = 0.85$ for OH only and OH with X, respectively. **b**, The same time series as presented in Fig. 1a, now showing, in addition to [H₂SO₄]_{meas} (black line) and calculated from OH reaction (blue area), the concentration calculated from the X reaction as in **a** (red area). Between 30 July and 2 August, the SO₂ concentration was mostly below the detection limit during which time the calculated concentration is not depicted. The error bars, not shown for clarity, are essentially the same as in Fig. 1a.

during some evenings and nights, addition of the new SO₂ oxidant significantly improves the overall agreement between measured and calculated H₂SO₄ concentrations.

The chemistry investigated here is tightly connected with the presence of biogenic volatile organic compounds (BVOCs), and thereby with forest emissions. Covering vast areas of the Earth's surface, forests play an important role in global cycles of carbon, water and energy⁶. BVOCs emitted by forests dominate the global secondary organic aerosol loading²⁷, and contribute significantly to the global budget of cloud condensation nuclei²⁸. Our findings add to the already substantial significance of forests in the Earth system by introducing a previously unknown oxidant, probably an sCI, capable of oxidizing at least SO₂ and possibly also other atmospheric trace gases relevant to atmospheric chemistry. Because gaseous sulphuric acid is formed in this process, the new chemistry is likely to affect the formation of new atmospheric particles, the production of secondary cloud condensation nuclei and ultimately climate. Our findings demonstrate a new connection between anthropogenic activities (SO₂ emissions), natural ecosystems (BVOC emissions and secondary organic aerosol formation) and climate (from cloud properties to radiative forcing). This connection is likely to change in the future as a result of changing SO₂ and BVOC emissions due to air quality regulations and warming climate²⁹. More detailed experimental and theoretical investigations are clearly needed to find out the importance of the new oxidant in atmospheric chemistry and climate at present and under future conditions.

METHODS SUMMARY

The results of the laboratory experiments described here were obtained from two different experimental systems. Both systems used the flow tube (that is, continuous flow) technique, where gases are added to a continuous stream and allowed to react for a known period of time. Sulphuric acid was then measured at the exit of the flow tube using nitrate-ion-based chemical ionization mass spectrometry (CIMS). One experimental apparatus (referred to as IfT-LFT) was located at the Leibniz-Institute for Tropospheric Research in Leipzig, Germany, and the other at the National Center for Atmospheric Research (NCAR) in Boulder, Colorado. The two systems differ from each other in their geometries, residence times and the method by which the reagent gases are introduced. The CIMS instrument was also used in the field measurements performed at the SMEAR II station located in a Finnish boreal forest for the detection of H₂SO₄, OH and X. A more detailed description of the methods is given in Supplementary Information.

Full Methods and any associated references are available in the online version of the paper.

Received 23 November 2011; accepted 10 May 2012.

- Liao, H. *et al.* Effect of chemistry-aerosol-climate coupling on predictions of future climate and future levels of tropospheric ozone and aerosols. *J. Geophys. Res.* **114**, D10306, <http://dx.doi.org/10.1029/2008JD010984> (2009).
- Rowland, F. S. Stratospheric ozone depletion. *Phil. Trans. R. Soc. Lond. B* **361**, 769–790 (2006).
- Likens, G. E., Bormann, F. H. & Johnson, N. M. Acid rain. *Environment* **14**, 33–40 (1974).
- Fenger, J. Air pollution in the last 50 years — from local to global. *Atmos. Environ.* **43**, 13–22 (2009).
- Lelieveld, J. *et al.* Atmospheric oxidation capacity sustained by a tropical forest. *Nature* **452**, 737–740 (2008).
- Arneth, A. *et al.* Terrestrial biogeochemical feedbacks in the climate system. *Nature Geosci.* **3**, 525–532 (2010).
- Su, H. *et al.* Soil nitrite as a source of atmospheric HONO and OH radicals. *Science* **333**, 1616–1618 (2011).
- Di Carlo, P. *et al.* Missing OH reactivity in a forest: evidence for unknown reactive biogenic VOCs. *Science* **304**, 722–725 (2004).
- Lou, S. *et al.* Atmospheric OH reactivities in the Pearl River Delta – China in summer 2006: measurements and model results. *Atmos. Chem. Phys.* **10**, 11243–11260 (2010).

- Hofzumahaus, A. *et al.* Amplified trace gas removal in the troposphere. *Science* **324**, 1702–1704 (2009).
- Montzka, S. A. *et al.* Small interannual variability of global atmospheric hydroxyl. *Science* **331**, 67–70 (2011).
- Sipilä, M. *et al.* The role of sulfuric acid in atmospheric nucleation. *Science* **327**, 1243–1246 (2010).
- Eisele, F. L. & Tanner, D. J. Ion-assisted tropospheric OH measurements. *J. Geophys. Res.* **96**, 9295–9308 (1991).
- Petäjä, T. *et al.* Sulfuric acid and OH concentrations in a boreal forest site. *Atmos. Chem. Phys.* **9**, 7435–7448 (2009).
- Hakola, H. Seasonal variation of VOC concentrations above a boreal coniferous forest. *Atmos. Environ.* **37**, 1623–1634 (2003).
- Lappalainen, H. K. *et al.* Day-time concentrations of biogenic volatile organic compounds in a boreal forest canopy and their relation to environmental and biological factors. *Atmos. Chem. Phys.* **9**, 5447–5459 (2009).
- Cocks, A. T., Fernando, R. P. & Fletcher, I. S. The gas-phase reaction of the methylperoxy radical with sulphur dioxide. *Atmos. Environ.* **20**, 2359–2366 (1986).
- Xie, Z. D. Formation mechanism of condensation nuclei in nighttime atmosphere and the kinetics of the SO₂-O₃-NO₂ system. *J. Phys. Chem.* **96**, 1543–1547 (1992).
- Kurtén, T., Lane, J. R., Jørgensen, S. & Kjaergaard, H. Nitrate radical addition-elimination reactions of atmospherically relevant sulfur-containing molecules. *Phys. Chem. Chem. Phys.* **12**, 12833–12839 (2010).
- Kurtén, T., Lane, J. R., Jørgensen, S. & Kjaergaard, H. A. Computational study of the oxidation of SO₂ to SO₃ by gas-phase organic oxidant. *J. Phys. Chem. A* **115**, 8669–8681 (2011).
- Hatakeyama, S., Kobayashi, H., Lin, Z.-Y., Tagaki, H. & Akimoto, H. Mechanism for the reaction of H₂COO with SO₂. *J. Phys. Chem.* **90**, 4131–4135 (1986).
- Johnson, D., Lewin, A. G. & Marston, G. The effect of Criegee-intermediate scavengers on the OH yield from the reaction of ozone with 2-methylbut-2-ene. *J. Phys. Chem. A* **105**, 2933–2935 (2001).
- Jiang, L., Xu, Y. & Ding, A. J. Reaction of stabilized Criegee intermediates from ozonolysis of limonene with sulfur dioxide: ab initio and DFT study. *J. Phys. Chem. A* **114**, 12452–1246 (2010).
- Welz, O. *et al.* Direct kinetic measurements of Criegee intermediate (CH₂OO) formed by reaction of CH₂I with O₂. *Science* **335**, 204–207 (2012).
- Cox, R. A. & Penkett, S. A. Oxidation of atmospheric SO₂ by products of the ozone-olefin reaction. *Nature* **230**, 321–322 (1971).
- Drozd, G. T., Kroll, J. & Donahue, N. M. 2,2-Dimethyl-2-butene (TME) ozonolysis: pressure dependence of stabilized Criegee intermediates and evidence of stabilized vinyl hydroperoxides. *J. Phys. Chem. A* **115**, 161–166 (2011).
- Heald, C. L. *et al.* Predicted change in global secondary organic aerosol concentrations in response to future climate, emissions, and land use change. *J. Geophys. Res.* **113**, D05211, <http://dx.doi.org/10.1029/2007JD009092> (2008).
- Spracklen, D. V. *et al.* Contribution of particle formation to global cloud condensation nuclei concentrations. *Geophys. Res. Lett.* **35**, L06808, <http://dx.doi.org/10.1029/2007GL033038> (2008).
- Arneth, A., Unger, N., Kulmala, M. & Andreae, M. O. Clean the air, heat the planet? *Science* **326**, 672–673 (2009).
- DeMore, W. *et al.* *Chemical Kinetics and Photochemical Data for Use in Stratospheric Modeling. Evaluation 12*. JPL Publication 97-4 (Jet Propulsion Laboratory, 1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

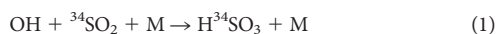
Acknowledgements We thank K. Pielok and A. Rohmer for technical assistance. This work was partially funded by the European Commission Sixth Framework programme project EUCAARI, contract no. 036833-2 (EUCAARI), the Academy of Finland (251427, 139656, Finnish centre of excellence 141135), the European Research Council (ATMNUCLE), the Kone Foundation, the Väisälä Foundation, the Maj and Tor Nessling Foundation (2010212), the Otto Malm Foundation and the US National Science Foundation.

Author Contributions R.L.M., T.B. and M.S. designed the experiments. R.L.M., T.B., M.S. and S.K. performed the laboratory experiments. R.L.M., T.P. and M.S. conducted the field measurements. T.B., T.K., and P.P. performed the model and theoretical calculations, and R.L.M., T.B., M.S. and P.P. analysed the data. All authors (R.L.M., T.B., M.S., P.P., T.P., S.K., T.K., F.S., V.-M.K., and M.K.) contributed to the interpretation and to manuscript preparation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.L.M. (roy.mauldin@helsinki.fi).

METHODS

CIMS measurements. Measurements of OH and H₂SO₄ were performed using the Chemical Ionization Mass Spectrometer (CIMS) technique. The technique has been described elsewhere^{13,31,32} therefore only details relevant to the present work will be discussed here. Briefly, sample air is drawn through the 1.9 cm stainless steel inlet, and a small amount ($\sim 10^{14}$ molecules cm⁻³) of isotopically labelled ³⁴SO₂ is added through a pair of 0.011 cm i.d. transversely opposed injectors located near the front opening. The OH is then converted into H₂³⁴SO₄ via the following reaction sequence:



Isotopically labelled SO₂ is used to discriminate between H₂SO₄ derived from OH and ambient H₂SO₄. To prevent cycling of HO₂ and RO₂ back into OH, propane is added on a continuous basis through a second pair of injectors located ~ 5 cm downstream of the first pair, after the ambient OH initially present has been converted into H₂SO₄. Propane is added through these injectors at sufficient concentrations to remove more than 99% of the OH which has been cycled back from HO₂ or RO₂ via reactions with NO or O₃. To account for other unknown processes which can convert SO₂ into H₂SO₄ an 'OH background' is performed, in which propane is added along with the ³⁴SO₂ through the front injectors at a concentration sufficient to remove >98% of the OH present. These OH background values are used in this work to describe the measurement of X. More details concerning the injectors and the sampling port chemistry are described elsewhere³¹. Once formed, the H₂³⁴SO₄ is measured in the same manner as H₂SO₄ via chemical ionization.

IFT-LFT. Experiments were carried out in the atmospheric pressure flow-tube IFT-LFT (i.d. 8 cm; length 505 cm) at 293 ± 0.5 K (ref. 33). The flow tube consists of a first section (56 cm) that includes the inlet system for gas input (air premixed with SO₂ from a calibration gas mixture (1 p.p.m.v. or 10 p.p.m.v. SO₂ in N₂ (Messer)), O₃ from an ozone generator outside the flow tube (UVP OG-2), the OH scavenger H₂ and the olefin premixed from a metering device). At the end of the tube, all sampling outlets are attached. O₃ and SO₂ concentrations were measured by means of gas monitors (Thermo Environmental Instruments: 49C and 43C) or by long-path ultraviolet absorption spectroscopy (Perkin-Elmer: Lambda 800) using a gas cell with a White-mirror optics adjusted at a path-length of 512 cm. The organics were followed by proton transfer reaction-mass spectrometry (PTR-MS) or by on-line gas chromatography-flame ionization detection (GC-FID) connected via a cryo-enrichment device. Sulphuric acid in the IFT-LFT was measured with a chemical ionization mass spectrometer, CIMS, in the same way as described for the NCAR experiments. The flow was set at 15 l min^{-1} (STP) resulting in a residence time of about 95 s.

H₂ (99.999%, Messer) was directly added to the carrier gas flow. As the carrier gas we used high-purity synthetic air (99.9999999%, Linde and further purification with GateKeeper CE-500KF-O-4R, Aeronex). All gas flows were set by means of calibrated gas flow controllers (MKS 1259/1179) and the pressure in the tube was measured using a capacitive manometer (Baratron).

NCAR. The reaction system consists of a glass flow tube with a movable stainless steel injector. A stream of hydrocarbon-free air (also called 'zero air' below)

containing SO₂ and the alkene being studied is added to the glass flow tube. Ozone is produced inside the stainless steel injector by passing a flow of O₂ over a mercury Pen-Ray lamp located inside the injector. The ozone is then introduced into the main flow at the end of the injector. The gases then can react as the flow proceeds towards the exit of the tube, where the flow is sampled. The reaction time can be varied by either adjusting the amount of the main flow, or by changing the position of the end of the injector.

The flow tube consists of a 71-cm-long Pyrex tube connected to a 20.3-cm-long Pyrex Y. Both pieces have 3.38 cm i.d. and are connected via no. 40 O-ring joints sealed with a silicone O-ring. The Pyrex Y allows access of the movable injector into the flow tube as well as providing a means to introduce the main flow. The injector consisted of a thin walled 100-cm-long, 1.27-cm-i.d. stainless steel tube with one end sealed and 24 0.2-mm holes drilled radially 0.5 cm from the sealed end. Inside the injector is a mercury Pen-Ray lamp. Ozone was produced by passing a flow of O₂ over the lamp. The lamp was situated such that the end of the lamp is ~ 5 cm from the sealed end to prevent radiation from the lamp photolysing the main flow. The injector was inserted into the main glass flow tube by means of a no. 40 O-ring joint reduced to a 1.9-cm o.d. tube, and sealed via a Swagelok fitting modified to use silicone O-rings.

The zero air used in this system was produced by filtering ambient air via a zero air generator (Adco). The UHP (ultra-high purity) oxygen was provided by General Air and had a stated purity of 99.9999%. The SO₂ used was a 0.5% mixture of SO₂ in UHP N₂ and was provided by Scott Speciality Gases. Alkene mixtures were made 'in house' at NCAR, and their concentrations determined via gas chromatography. All flows into the flow tube were controlled by means of mass flow controllers (MKS).

The flow was sampled at the exit of the flow tube by a CIMS^{13,14} (Chemical Ionization Mass Spectrometer) measuring H₂SO₄, a PTR-MS measuring various hydrocarbon products, and an O₃ analyser (2B Technologies).

Field measurements. Field measurements were conducted at the SMEAR II field station in Finnish boreal forest³⁴. The station (61° 51' N, 24° 17' E) is situated in southern Finland about 60 km northeast of the city of Tampere. The nearest village with some industrial activity is approximately 10 km away, and the nearest buildings are by a small lake 500 m away from the measurement station. The station is surrounded by a coniferous Scots pine dominated forest. Other major species include spruce and birch. All field measurements discussed in this Letter were done in a container located in a small open area surrounded by the forest. The SMEAR II station is equipped with extensive meteorological and gas and aerosol instrumentation.

The calculated concentration of sulphuric acid resulting from the reaction of SO₂ with either OH or X was obtained by assuming a steady state between the sulphuric acid production and its loss by condensation onto pre-existing aerosol particles. Detailed descriptions of calculations are given in Supplementary Information

31. Tanner, D. J., Jefferson, A. & Eisele, F. L. Selected ion chemical ionization mass spectrometric measurement of OH. *J. Geophys. Res.* **102**, 6415–6425 (1997).
32. Mauldin, R. L. III *et al.* OH measurements during ACE-1: observations and model comparisons. *J. Geophys. Res.* **103**, 16713–16729 (1998).
33. Berndt, T. *et al.* Laboratory study on new particle formation from the reaction OH + SO₂: influence of experimental conditions, H₂O vapour, NH₃ and the amine tert-butylamine on the overall process. *Atmos. Chem. Phys.* **10**, 7101–7116 (2010).
34. Hari, P. & Kulmala, M. Station for measuring ecosystem atmosphere relations (SMEAR II). *Boreal Environ. Res.* **10**, 315–322 (2005).

Water balance of global aquifers revealed by groundwater footprint

Tom Gleeson¹, Yoshihide Wada², Marc F. P. Bierkens^{2,3} & Ludovicus P. H. van Beek²

Groundwater is a life-sustaining resource that supplies water to billions of people, plays a central part in irrigated agriculture and influences the health of many ecosystems^{1,2}. Most assessments of global water resources have focused on surface water^{3–6}, but unsustainable depletion of groundwater has recently been documented on both regional^{7,8} and global scales^{9–11}. It remains unclear how the rate of global groundwater depletion compares to the rate of natural renewal and the supply needed to support ecosystems. Here we define the groundwater footprint (the area required to sustain groundwater use and groundwater-dependent ecosystem services) and show that humans are overexploiting groundwater in many large aquifers that are critical to agriculture, especially in Asia and North America. We estimate that the size of the global groundwater footprint is currently about 3.5 times the actual area of aquifers and that about 1.7 billion people live in areas where groundwater resources and/or groundwater-dependent ecosystems are under threat. That said, 80 per cent of aquifers have a groundwater footprint that is less than their area, meaning that the net global value is driven by a few heavily overexploited aquifers. The groundwater footprint is the first tool suitable for consistently evaluating the use, renewal and ecosystem requirements of groundwater at an aquifer scale. It can be combined with the water footprint and virtual water calculations^{12–14}, and be used to assess the potential for increasing agricultural yields with renewable groundwater¹⁵. The method could be modified to evaluate other resources with renewal rates that are slow and spatially heterogeneous, such as fisheries, forestry or soil.

The ecological footprint and the water footprint are powerful, popular and complementary tools for planning, education and public awareness, but their methodologies are fundamentally different¹³. The ecological footprint is the land area (in km²) required to sustain a population¹⁶, whereas the water footprint is the volume (in m³ yr^{−1}) of freshwater required¹². The ecological footprint directly defines the ecological impact of human consumption by comparing the available bioproductive area to the area required for the consumption of specific goods and services. The water footprint tracks the volume of virtual water used by a population, where virtual water is the volume of freshwater used to produce a commodity, good or service along the various steps of production^{13,14,17}. The water footprint quantifies the components of virtual water: green water (soil water), blue water (surface water and groundwater) and grey water (polluted water). However, until recently¹⁸ the water footprint was not able to assess the impact of our water consumption on natural stocks and flows¹⁹ because it generally focused on the volumes of water required without quantifying the volume of water available. The groundwater footprint, as proposed here, is complementary to the well-established water footprint method and can be used to assess the impact of our groundwater consumption on natural stocks and flows. Here, we apply the groundwater footprint methodology globally to regional-scale, hydrologically active aquifers²⁰ (see Supplementary Information for a definition). The focus of the method is currently on groundwater quantity rather than

quality, which is a conservative assumption that results in smaller groundwater footprints for aquifers affected by groundwater contamination. The groundwater footprint method can be applied to a variety of scales and contexts like the ecological footprint, water footprint and virtual water concepts^{12–14}.

We define the groundwater footprint as the area required to sustain groundwater use and groundwater-dependent ecosystem services of a region of interest, such as an aquifer, watershed or community. The groundwater footprint (GF) is defined more formally as $GF = A[C/(R - E)]$, where C , R and E are respectively the area-averaged annual abstraction of groundwater, recharge rate, and the groundwater contribution to environmental streamflow, all in units with dimensions of length/time, such as m d^{−1} (Supplementary Fig. 1). A (in units of length², such as m²) is the areal extent of any region of interest where C , R and E can be defined. The groundwater footprint is essentially a water balance between aquifer inflows (R) and outflows (C and E) which can be derived from observations and/or model output. C is derived directly from the use of groundwater at the scale of interest although actual groundwater abstraction is often poorly known^{9,11}. R is the long-term natural areal flux into the system plus the additional recharge from irrigation, and can be derived from geochemical tracer methods or hydrologic models^{21,22}. E is the quantity of groundwater that needs to be allocated to surface water flow to sustain ecosystem services, which is most important during low flow conditions^{23,24}. Thus, the groundwater footprint method emphasizes the contribution of groundwater to the environmental requirements during low flows, although natural streamflow variability is also essential to maintaining the environmental integrity of surface water systems²⁵. Environmental flow requirements for specific aquifers or watersheds are most accurately determined by detailed hydroecological data and multidisciplinary expert consultation at the scale of specific aquifers or watersheds^{24,25}. The Supplementary Information contains the mathematical relationship of the groundwater footprint to the ecological footprint¹⁶ and previous water stress indicators^{4–6,24}, as well as other forms of the groundwater footprint equation that may be useful for local calculations with different data sources.

We calculated the global groundwater footprint as the sum of the groundwater footprints of large aquifers worldwide using spatially distributed recharge rates and environmental flows derived from PCR-GLOBWB and gridded groundwater consumption estimates²⁶ (see Methods Summary and Fig. 1). PCR-GLOBWB is a conceptual, process-based global hydrologic model that simulates the daily water balance for 1958–2000 at 0.5° resolution (that is, ~50 km at the Equator) and is validated to GRACE satellite observations and global streamflow estimates^{10,22,26}. Recharge (R) is the long-term natural groundwater recharge and additional recharge from irrigation derived from ref. 10. For global-scale assessment of water resources, low flow requirements based on consistent hydrologic criteria are useful¹⁷. Therefore, following ref. 24, environmental flows (E) were taken to be equal to Q_{90} , the monthly streamflow that is exceeded 90% of the time during the period 1958–2000. We calculated environmental flows

¹Department of Civil Engineering, McGill University, Montreal, Quebec H3A 0C3, Canada. ²Department of Physical Geography, Faculty of Geosciences, Utrecht University, PO Box 80115, 3508 TC Utrecht, The Netherlands. ³Deltares, PO Box 85467, 3508 AL Utrecht, The Netherlands.

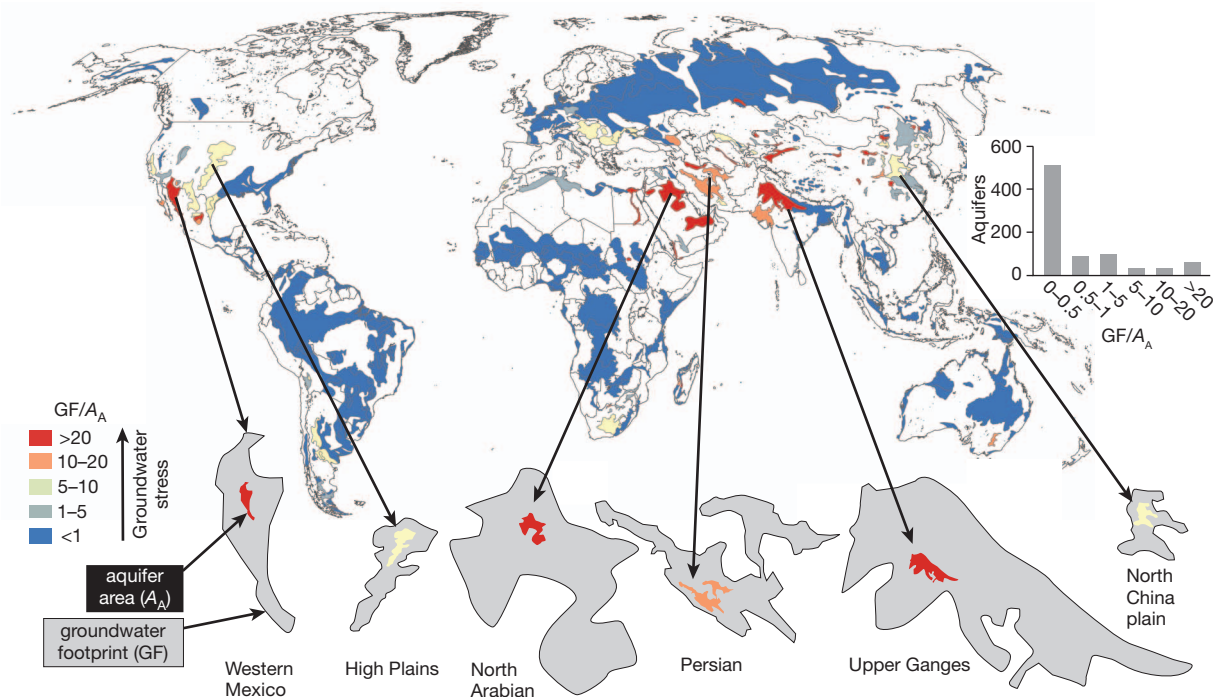


Figure 1 | Groundwater footprints of aquifers that are important to agriculture are significantly larger than their geographic areas. Aquifers are major groundwater basins with recharge of $>2 \text{ mm yr}^{-1}$ in the global inventory of groundwater resources²⁰ (see Supplementary Information). At the bottom of the figure, the areas of the six aquifers (Western Mexico, High Plains, North

Arabian, Persian, Upper Ganges and North China plain) are shown at the same scale as the global map; the surrounding grey areas indicate the groundwater footprint proportionally at the same scale. The ratio GF/A_A indicates widespread stress of groundwater resources and/or groundwater-dependent ecosystems. Inset, histogram showing that GF is less than A_A for most aquifers.

at the basin scale and the associated groundwater requirement expressed as a uniform fraction of total recharge (see Methods Summary and Supplementary Fig. 2). This basin-scale fraction was then multiplied with the grid-based recharge to obtain E . We derived grid-based groundwater abstraction, C , for the year 2000 (ref. 26; see Supplementary Information) and subsequently aggregated the fluxes C , R and E over hydrologically active regional aquifers²⁰ to calculate their groundwater footprint. Aggregation at the scale of regional aquifers is justifiable given the resolution of the input data while naturally integrating lateral groundwater flow that might occur due to abstraction wells.

Figure 1 is, to our knowledge, the first spatially explicit comparison of groundwater use, availability and environmental flow for aquifers globally. We acknowledge that each of the regional aquifers has significant internal heterogeneity and that groundwater extractions

often acutely affect smaller regions within aquifers, although we partly account for this heterogeneity by using the highest available resolution of regional aquifers²⁰ as the basis for aggregation. A few aquifers with well-documented histories of groundwater depletion have large groundwater footprints (for example, the Upper Ganges, High Plains, North China plain and Central Valley^{7,8,27,28}; Table 1). A number of other aquifers with large groundwater footprints (for example, the Persian, Arabian and Western Mexico aquifers) are not as well documented, although evidence of groundwater depletion in these aquifers is discussed in non-peer-reviewed literature (see Supplementary Information). It is instructive to compare the ratio of groundwater footprint (GF) to aquifer area (A_A), which is a groundwater stress indicator (see Supplementary Information). $\text{GF}/A_A > 1$ indicates where unsustainable groundwater consumption could affect groundwater availability and groundwater-dependent surface water

Table 1 | Properties of aquifers with the largest groundwater footprints

Aquifer	Country	GF (10^6 km^2)	A_A (10^6 km^2)	GF/A_A
Upper Ganges	India, Pakistan	26.1 ± 7.5	0.48	54.2 ± 15.6
North Arabian	Saudi Arabia	17.3 ± 4.7	0.36	48.3 ± 13.5
South Arabian	Saudi Arabia	9.5 ± 3.6	0.25	38.5 ± 14.7
Persian	Iran	8.4 ± 3.7	0.42	19.7 ± 8.6
South Caspian	Iran	5.9 ± 2.0	0.06	98.3 ± 32.6
Western Mexico	Mexico	5.5 ± 2.0	0.21	26.6 ± 9.4
High Plains	USA	4.5 ± 1.2	0.50	9.0 ± 2.4
Lower Indus	India, Pakistan	4.2 ± 1.5	0.23	18.4 ± 6.5
Nile delta	Egypt	3.1 ± 0.8	0.10	31.7 ± 7.9
Danube basin	Hungary, Austria, Romania	2.4 ± 0.8	0.32	7.4 ± 2.6
Central Mexico	Mexico	1.8 ± 0.5	0.20	9.1 ± 2.6
North China plain	China	1.8 ± 0.6	0.23	7.9 ± 2.8
Northern China	China	1.4 ± 0.6	0.31	4.5 ± 1.8
North Africa	Algeria, Tunisia, Libya	0.9 ± 0.3	0.36	2.6 ± 0.9
Central Valley	USA	0.4 ± 0.2	0.07	6.4 ± 2.4
Other aquifers		38.6 ± 10.8	34.17	1.1 ± 0.3
All aquifers		131.8 ± 24.9	38.27	3.5 ± 0.7

The values of GF (groundwater footprint) and GF/A_A are the mean and standard deviation of 10,000 Monte Carlo realizations based on independent estimates of recharge and abstraction¹⁰. Note that only the 15 aquifers with the largest GF are listed individually. The remaining 768 'other aquifers' are included in 'all aquifers'. GF/A_A is calculated before rounding the GF to one decimal place. A_A is aquifer area.

and ecosystems. The ratio GF/A_A is $\gg 1$ for the aquifers with large groundwater footprints mentioned above, indicating unsustainable groundwater mining, often of fossil groundwater recharged under past climatic conditions (Table 1). However, the majority of aquifers in the world have $GF < 10^6 \text{ km}^2$ and 80% of aquifers have $GF/A_A < 1$, suggesting that groundwater depletion is not ubiquitous (Fig. 1).

The size of the global groundwater footprint is currently $(131.8 \pm 24.9) \times 10^6 \text{ km}^2$, or 3.5 ± 0.7 times the actual area of hydrologically active aquifers²⁰. Even if no groundwater is allocated for environmental flows ($E = 0$), the global groundwater footprint is still $(76.5 \pm 15.7) \times 10^6 \text{ km}^2$ or 2.0 ± 0.4 times the actual aquifer area (Supplementary Table 2). The global groundwater footprint is dominated by a handful of countries, including the United States, China, Pakistan, Iran, India, Mexico and Saudi Arabia (Table 1). The ratio of global groundwater consumption to the difference between global recharge and global environmental streamflow is ~ 0.2 . High recharge rates in some ecologically sensitive areas, such as the Amazon, are included in this calculation but practically cannot be used to balance overexploitation in arid regions. 1.7 ± 0.4 billion people live in regions with $GF/A_A > 1$, where groundwater consumption could affect groundwater availability and/or groundwater-dependent surface water and ecosystems in the future. Approximately 60% of the people²⁹ living in regions with $GF/A_A > 1$ are located in India and China (Fig. 2a).

The groundwater footprint can be used to assess the potential to increase agricultural yields with renewable groundwater, or can be combined with water footprint and virtual water calculations. Foley *et al.*¹⁵ calculated the global distribution of potential new calories that could be derived by bringing the world's agricultural yields to within 95% of their potential for 16 major crops. Crop yields may be limited by a number of factors, including water or nutrient availability and management¹⁵, but increasing agricultural yields generally leads to increased water demand. Because groundwater is critical for irrigation in many agricultural regions, it is useful to assess how the spatial distribution of groundwater stress (Fig. 1) compares to the potential for new calories (figure 3 in ref. 15). Figure 2b shows that some areas with potential new calories coincide with aquifers that are less stressed, suggesting there is potential that renewable groundwater could be used sustainably to increase crop yields. However, aquifers that are significantly stressed ($GF/A_A \gg 1$) also underlie areas with potential new calories, and in these regions groundwater cannot be used sustainably to increase yields. These analyses only consider the groundwater footprint and agricultural yields, and should be placed in a broader socio-economic context.

We now show how the groundwater footprint can be used to assess the impact of transferring groundwater consumption between regions: the Upper Ganges aquifer in northwestern India and Pakistan has the largest groundwater footprint and a large GF/A_A ratio (Table 1), but the Lower Ganges aquifer has a GF/A_A ratio of less than one owing to low groundwater consumption and high recharge rates (Supplementary Fig. 3). Transferring even a small percentage of the groundwater consumption of the Upper Ganges to the Lower Ganges leads to a significant decrease in the combined groundwater footprint and GF/A_A ratio, because the aquifer-scale recharge rate to the Lower Ganges aquifer is approximately ten times higher than to the Upper Ganges aquifer (Supplementary Fig. 3). However, even if all the groundwater consumption of the Upper Ganges is transferred, the combined GF/A_A ratio remains greater than one, indicating that the current groundwater consumption in the region cannot be made sustainable by transferring groundwater consumption. In the future, the groundwater footprint could potentially serve as a metric to assess to what extent renewable groundwater could be exploited in virtual water trade schemes.

We stress that all variables used in our calculations, except for area, are subject to uncertainty. Least known are the environmental flow requirements, which are dependent on expert consultation and

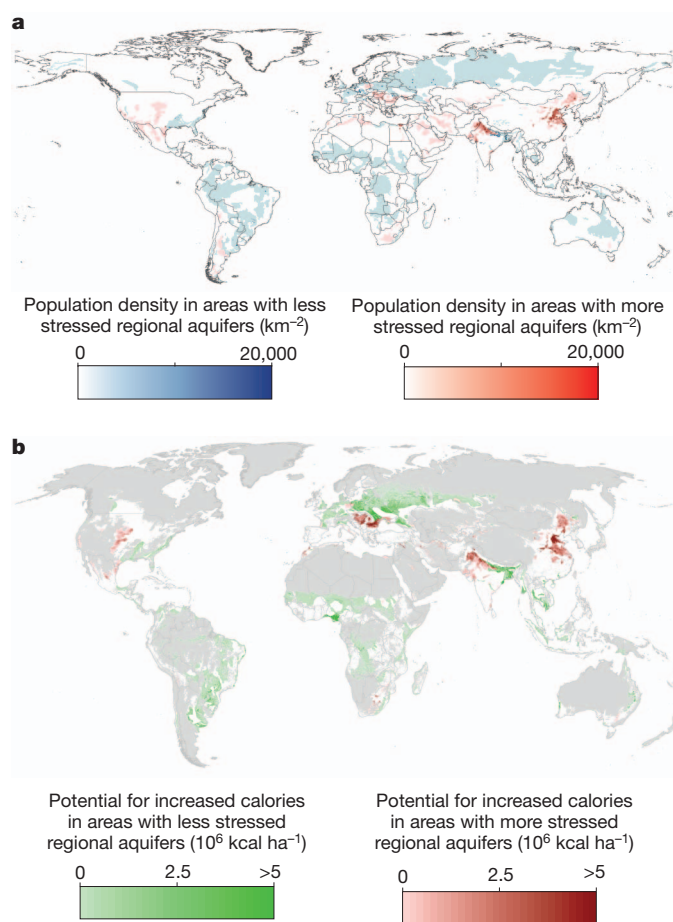


Figure 2 | Groundwater stress may be affecting ~ 1.7 billion people and could limit the potential to increase agricultural production. The ratio GF/A_A is used to differentiate areas with less groundwater stress ($GF/A_A < 1$) and more groundwater stress ($GF/A_A > 1$). **a**, Population densities, derived from the gridded population of the world for year 2000 (ref. 29). Areas that do not have underlying regional aquifers, or that have very low population density are shown in white. **b**, Potential for increased calories (see main text). Some areas with potential new calories¹⁵ coincide with stressed aquifers and some areas coincide with aquifers that are less stressed. Areas with potential new calories that are not underlain by a regional aquifer are shown in white.

detailed hydroecological relationships that are often poorly defined. However, in our global calculations we kept the environmental flow conditions constant, as it is primarily a management decision at regional to national scales and we explicitly incorporated the uncertainty due to recharge and groundwater consumption using a Monte Carlo analysis with 10,000 realizations (see Methods Summary). In the future, other data sets, including sub-national groundwater consumption data (Supplementary Fig. 4), could be used, where available, to calculate the groundwater footprint at different scales or for different administrative units.

The groundwater footprint is a powerful and hydrologically grounded tool for groundwater analysis and policy that complements and extends the ecological footprint, water footprint and virtual water methods. It is an advance on previous work on groundwater depletion^{6–8} as it explicitly includes environmental flows, it considers aquifers as a hydrologically grounded scale of analysis, it is more intuitive to water managers and the general public than depletion volumes, and it is based on improved estimates of recharge and abstraction. As exemplified above, the groundwater footprint refocuses the discussion to solutions, making it a valuable water management and policy tool. Practically, it allows short-term water resource monitoring and management measures to focus on the handful of aquifers with egregious groundwater footprints rather than dissipating

efforts across all aquifers. Additionally, the groundwater footprint can be used to assess the potential to achieve increased agricultural yields with sustainable groundwater. Also, as satellite-based groundwater depletion data sets (GRACE) are emerging^{7,8}, the groundwater footprint offers a useful framework for analysing these global depletion data sets in a broader framework of groundwater resource use, availability and environmental flows. Last, because the groundwater footprint method is flexible and spatially distributed, it could be modified for other resources whose renewal is slow and spatially heterogeneous, such as fisheries, forestry or soil.

METHODS SUMMARY

Global, spatially distributed estimates of the recharge rate, R , were obtained from the hydrological model PCR-GLOBWB²² using a global permeability map³⁰. Annual fields were averaged⁹ to obtain the average recharge rate over the period 1958–2000 with a spatial resolution of 0.5°. Artificial recharge due to irrigation water was added¹⁰. In the absence of global information on environmental flow requirements, the monthly streamflow exceeded 90% of the time (Q_{90}) was adopted²². Per basin, Q_{90} was determined at the basin outlet for the entire simulation period. A basin-wide uniform fraction was computed by which the groundwater recharge contributes to the environmental flow requirement E . Grid-based annual groundwater abstraction, C , was derived from reported country statistics for the year 2000 (<http://www.un-igrac.org>), which were downscaled spatially relative to the local surface water deficit or total water demand depending on the situation per country²⁶. The above quantities of R , E and C were aggregated over hydrologically active, regional aquifers²⁰ to compute the aquifer-scale groundwater footprint. To account for uncertainty, we made use of the uncertainty estimates for recharge and groundwater abstraction of ref. 9. E was excluded from the uncertainty analysis, as it is often defined a priori as a management decision at regional to national scales. Following ref. 9, a Monte Carlo simulation with 100 independent realizations of R and C returned 10,000 values for the groundwater footprint for the hydrologically active aquifers, from which the mean and standard deviation were computed. The groundwater footprints were (1) summed globally to compare to the actual aquifer area, (2) used to calculate the affected population numbers using the gridded global population for year 2000 (ref. 29) and (3) compared to the spatial distribution of potential new calories¹⁵. See Supplementary Information for details on methods and data sets, as well for additional validation, including regional groundwater abstraction (Supplementary Fig. 4 and Supplementary Table 3).

Received 28 February; accepted 1 June 2012.

- Giordano, M. Global groundwater? Issues and solutions. *Annu. Rev. Environ. Resour.* **34**, 153–178 (2009).
- Siebert, S. *et al.* Groundwater use for irrigation — a global inventory. *Hydrol. Earth Syst. Sci.* **14**, 1863–1880 (2010).
- Postel, S. L., Daily, G. C. & Ehrlich, P. R. Human appropriation of renewable fresh water. *Science* **271**, 785–788 (1996).
- Vorosmarty, C. J., Green, P., Salisbury, J. & Lammers, R. B. Global water resources: vulnerability from climate change and population growth. *Science* **289**, 284–288 (2000).
- Alcamo, J. *et al.* Global estimates of water withdrawals and availability under current and future business-as-usual conditions. *Hydrol. Sci. J.* **48**, 339–348 (2002).
- Oki, T. & Kanae, S. Global hydrological cycles and world water resources. *Science* **313**, 1068–1072 (2006).
- Rodell, M., Velicogna, I. & Famiglietti, J. S. Satellite-based estimates of groundwater depletion in India. *Nature* **460**, 999–1002 (2009).
- Famiglietti, J. S. *et al.* Satellites measure recent rates of groundwater depletion in California's Central Valley. *Geophys. Res. Lett.* **38**, L03403, <http://dx.doi.org/10.1029/2010GL046442> (2011).
- Wada, Y. *et al.* Global depletion of groundwater resources. *Geophys. Res. Lett.* **37**, L20402, <http://dx.doi.org/10.1029/2010GL044571> (2010).
- Wada, Y., van Beek, L. P. H. & Bierkens, M. F. P. Nonsustainable groundwater sustaining irrigation: a global assessment. *Wat. Resour. Res.* **48**, W00L06, <http://dx.doi.org/10.1029/2011WR010562> (2012).
- Konikow, L. F. Contribution of global groundwater depletion since 1900 to sea-level rise. *Geophys. Res. Lett.* **38**, L17401, <http://dx.doi.org/10.1029/2011GL048604> (2011).
- Hoekstra, A. Y., Chapagain, A. K., Aldaya, M. M. & Mekonnen, M. M. *The Water Footprint Assessment Manual: Setting the Global Standard* (Earthscan, 2011).
- Hoekstra, A. Y. Human appropriation of natural capital: a comparison of ecological footprint and water footprint analysis. *Ecol. Econ.* **68**, 1963–1974 (2009).
- Allan, J. A. Virtual water: a strategic resource, global solutions to regional deficits. *Ground Water* **36**, 545–546 (1998).
- Foley, J. A. *et al.* Solutions for a cultivated planet. *Nature* **478**, 337–342 (2011).
- Wackernagel, M. & Rees, W. *Our Ecological Footprint* (New Society Publishers, 1996).
- Hoekstra, A. Y. & Mekonnen, M. M. The water footprint of humanity. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1109936109> (published online, 13 February 2012).
- Hoekstra, A. Y., Mekonnen, M. M., Chapagain, A. K., Mathews, R. E. & Richter, B. D. Global monthly water scarcity: blue water footprints versus blue water availability. *PLoS ONE* **7**, e32688 (2012).
- Hoekstra, A. Y., Gerbens-Leenes, W. & van der Meer, T. H. Reply to Pfister and Hellweg: Water footprint accounting, impact assessment, and life-cycle assessment. *Proc. Natl Acad. Sci. USA* **106**, E114 (2009).
- BGR/UNESCO. Groundwater Resources of the World 1 : 25 000 000 http://www.whymap.org/whymap/EN/Products/products_node_en.html (2008).
- Döll, P. Vulnerability to the impact of climate change on renewable groundwater resources: a global-scale assessment. *Environ. Res. Lett.* **4**, 035006 (2009).
- van Beek, L. P. H., Wada, Y. & Bierkens, M. F. P. Global monthly water stress: 1. Water balance and water availability. *Wat. Resour. Res.* **47**, W07517 (2011).
- Smakhtin, V. U. Low flow hydrology: a review. *J. Hydrol.* **240**, 147–186 (2001).
- Smakhtin, V. U., Revenga, C. & Döll, P. A pilot global assessment of environmental water requirements and scarcity. *Wat. Int.* **29**, 307–317 (2004).
- Poff, N. L. *et al.* The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards. *Freshwat. Biol.* **55**, 147–170 (2010).
- Wada, Y. *et al.* Global monthly water stress: 2. Water demand and severity of water stress. *Wat. Resour. Res.* **47**, W07518 (2011).
- Sophocleous, M. Review: groundwater management practices, challenges, and innovations in the High Plains aquifer, USA — lessons and recommended actions. *Hydrogeol. J.* **18**, 559–575 (2010).
- Foster, S. *et al.* Quaternary Aquifer of the North China Plain—assessing and achieving groundwater resource sustainability. *Hydrogeol. J.* **12**, 81–93 (2004).
- CIESIN. *Gridded Population of the World Version 3 (GPWv3): Population Density Grids* (Socioeconomic Data and Applications Center (SEDAC), Columbia University, 2011); available at <http://sedac.ciesin.columbia.edu/gpw> (accessed, 17 February 2011).
- Gleeson, T. *et al.* Mapping permeability over the surface of the Earth. *Geophys. Res. Lett.* **38**, L02401, <http://dx.doi.org/10.1029/2010GL045565> (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements S. Siebert, M. Jellinek, M. Lathuilliere, A. Henderson and W. Rees read or discussed earlier versions of the manuscript, which markedly improved it. T.G. was supported by the Natural Sciences and Engineering Research Council of Canada and a Canadian Institute for Advanced Research junior fellowship. Y.W. was supported by Utrecht University Focus Areas Theme 'Earth and sustainability'.

Author Contributions T.G. developed the groundwater footprint method, created the figures and wrote the paper with input from all authors. Y.W. and L.P.H.v.B. completed the analysis of groundwater consumption and hydrologic data. L.P.H.v.B., T.G. and M.F.P.B. developed the environmental flow methodology. All authors discussed results and edited the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.G. (tom.gleeson@mcgill.ca).

New fossils from Koobi Fora in northern Kenya confirm taxonomic diversity in early *Homo*

Meave G. Leakey^{1,2}, Fred Spoor^{3,4}, M. Christopher Dean⁴, Craig S. Feibel⁵, Susan C. Antón⁶, Christopher Kiarie¹ & Louise N. Leakey^{1,2}

Since its discovery in 1972 (ref. 1), the cranium KNM-ER 1470 has been at the centre of the debate over the number of species of early *Homo* present in the early Pleistocene epoch² of eastern Africa. KNM-ER 1470 stands out among other specimens attributed to early *Homo* because of its larger size, and its flat and subnasally orthognathic face with anteriorly placed maxillary zygomatic roots³. This singular morphology and the incomplete preservation of the fossil have led to different views as to whether KNM-ER 1470 can be accommodated within a single species of early *Homo* that is highly variable because of sexual, geographical and temporal factors^{4–9}, or whether it provides evidence of species diversity marked by differences in cranial size and facial or masticatory adaptation^{3,10–20}. Here we report on three newly discovered fossils, aged between 1.78 and 1.95 million years (Myr) old, that clarify the anatomy and taxonomic status of KNM-ER 1470. KNM-ER 62000, a well-preserved face of a late juvenile hominin, closely resembles KNM-ER 1470 but is notably smaller. It preserves previously unknown morphology, including moderately sized, mesiodistally long postcanine teeth. The nearly complete mandible KNM-ER 60000 and mandibular fragment KNM-ER 62003 have a dental arcade that is short anteroposteriorly and flat across the front, with small incisors; these features are consistent with the arcade morphology of KNM-ER 1470 and KNM-ER 62000. The new fossils confirm the presence of two contemporary species of early *Homo*, in addition to *Homo erectus*, in the early Pleistocene of eastern Africa.

KNM-ER 62000, with an estimated geological age of 1.91 to 1.95 Myr old (Fig. 1; metrics in Supplementary Note 1), consists of the maxillae, the palatine bones and the right zygomatic bone of a late juvenile. On the left, the lateral half of the maxillary body, the vertical plate of the palatine bone and the temporal process of the zygomatic bone are

missing. The lateral half of the right infra-orbital area is weathered, and the left alveolar process shows some medial displacement and damage to the lateral surface. The preserved teeth include the right P⁴, both M¹s and M²s, and the unerupted crown of the right M³, exposed in its crypt above the M². The occlusal surface of the right M¹ is missing, whereas the left M¹ has a partially damaged cervical margin. In its dental development, KNM-ER 62000 most closely resembles a 13- or 14-year-old modern human. But in the hominin fossil record, at a time when development was quicker, a close match is KNM-WT 15000, whose chronological age has been estimated at approximately 8 years (ref. 21; Supplementary Note 2).

In facial size, KNM-ER 62000 is similar to smaller specimens attributed to early *Homo* and *H. erectus* (for example, KNM-ER 1813 and D2700). The preserved tooth crowns are much smaller than in *Paranthropus boisei*, larger than in eastern African *H. erectus* and most similar in size to early *Homo* (Supplementary Notes 1 and 3). In shape they stand out by being relatively long mesiodistally, with the molars being rhomboidal in outline. The P⁴ is two-rooted with an incompletely divided buccal root; the P³ alveoli indicate a similar morphology. The predicted adult palate shape is relatively shorter than in australopiths (Supplementary Note 4). The alveolar process has parallel postcanine rows, showing the greatest breadth lingually at P⁴ and buccally at M¹.

The facial morphology of KNM-ER 62000 is derived compared with *Australopithecus* and *Paranthropus*, and its affinities are with the genus *Homo*. It shows striking similarity to KNM-ER 1470 (Supplementary Note 5), sharing many of the features that single out the latter as unique among specimens of early *Homo*³. A well-defined P³ jugum marks the 'corner' between the lateral and anterior surfaces of the maxillary alveolar process, so that the canine alveoli are fully part of the anterior row. The incisor alveoli are narrow mesiodistally for early



Figure 1 | The KNM-ER 62000 face. a–d, Anterior (a), right lateral (b), inferior (c) and superior views (d) of the KNM-ER 62000 face. Scale bar, 3 cm.

¹Turkana Basin Institute, PO Box 24926, Nairobi 00502, Kenya. ²Department of Anthropology, Stony Brook University, Stony Brook, New York 11794, USA. ³Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany. ⁴Department of Cell and Developmental Biology, University College London, London WC1E 6BT, UK. ⁵Department of Earth and Planetary Sciences, and Department of Anthropology, Rutgers University, Piscataway, New Jersey 08854-8066, USA. ⁶Department of Anthropology, New York University, New York, New York 10003, USA.



Figure 2 | The KNM-ER 60000 mandible and KNM-ER 62003 mandible fragment. a–d, Left lateral (a), right lateral (b), anterior (c) and occlusal (d) views of KNM-ER 60000. e, Occlusal view of KNM-ER 62003, with lines marking the mid-sagittal plane aligned to that of KNM-ER 60000. Scale bar, 3 cm.

Homo, show little anterior projection beyond the bicanine line, and the overlying subnasal area is transversely flat (Fig. 1c–d). In lateral view, a particularly straight facial profile is marked by a highly orthognathic nasoalveolar clivus (Fig. 1b; Supplementary Notes 5e and 6b, e). There is a distinct nasal sill, and the lateral nasal margins are rounded inferiorly, but mildly everted superiorly, implying some projection of the nasal bridge. The anterior surface of the zygomatic process is positioned between P^3 and P^4 . The midface is transversely flat with an infra-orbital area that slopes antero-inferiorly. The infraorbital foramen is placed medially and close to the orbital margin. The preserved inter-orbital morphology indicates that the nasal cavity was tall. Bi-orbital to maxillo-alveolar breadth proportions indicate that the upper face is distinctly narrow (Supplementary Note 1).

KNM-ER 60000 is a nearly complete adult mandible with an estimated geological age of 1.78 to 1.87 Myr old (Fig. 2; metrics in Supplementary Note 7). The well-preserved left side lacks only the anterior part of the ramus, including the coronoid process, but this area is well-preserved on the right side, facilitating full reconstruction (Supplementary Note 8). The right side of the corpus shows post-mortem deformation. The well-worn dentition with fully formed roots is complete through M_3 . The anterior arcade is flat across the front because narrow and short-rooted incisors sit medial, but barely anterior to the canines. The labiolingual incisor crown dimensions fall below the range observed for most eastern African *Homo* and *Paranthropus*. The P_3 to M_2 crowns are small for *Paranthropus*, and their sizes are below or at the low end of the size range for early *Homo*,

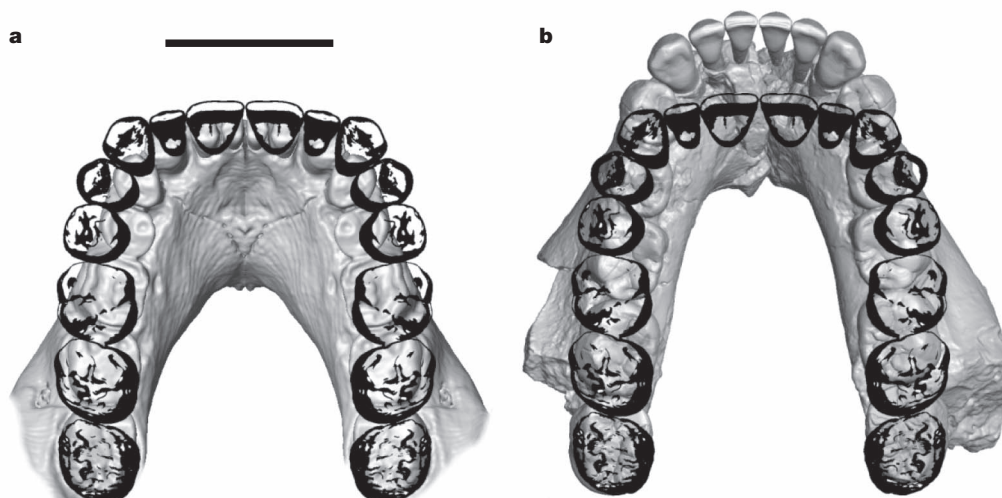
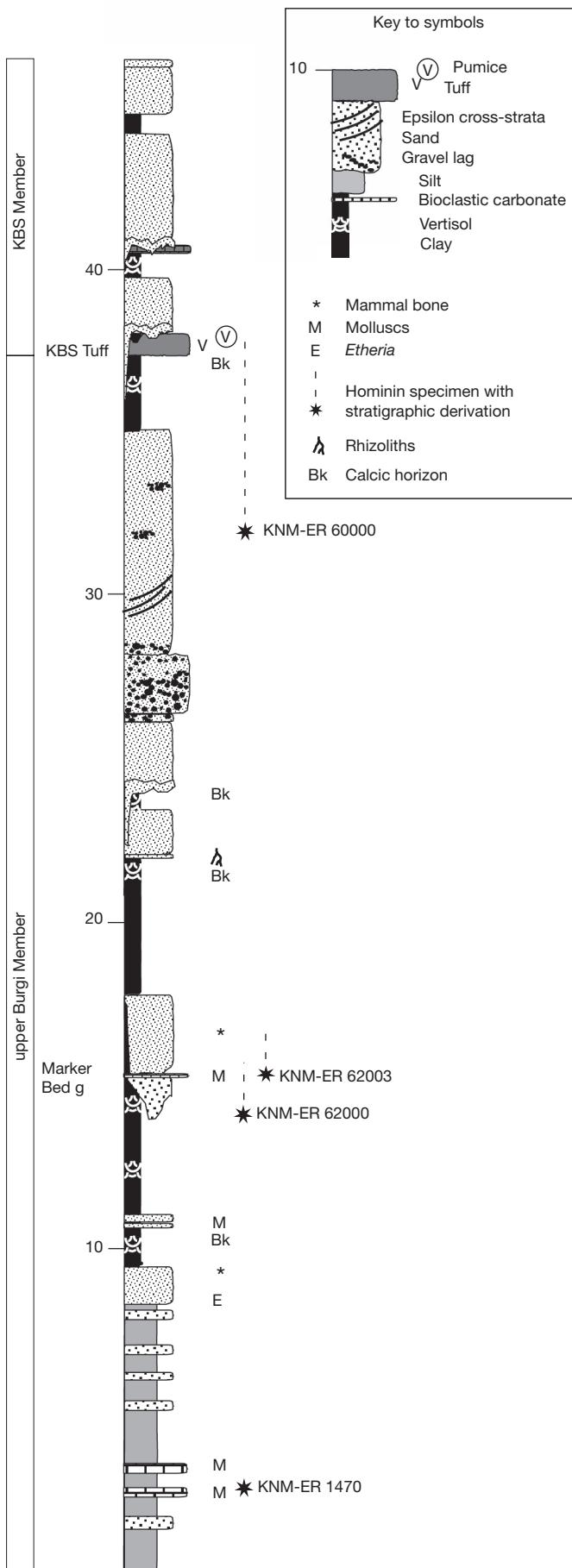


Figure 3 | Dental arcades compared. The reconstructed upper arcade of KNM-ER 62000 (outlined in black) occluded with the reconstructed lower arcades of KNM-ER 60000 (a) and KNM-ER 1802 (b), showing a good match with KNM-ER 60000 but not with KNM-ER 1802. Scale bar, 3 cm. See Supplementary Note 10 for details.



particularly their small mesiodistal dimensions (Supplementary Note 9). However, the M_3 dimensions are well within the range of early *Homo*, and the molars increase in size from M_1 to M_3 . Both M_2 and M_3 show a C7 cusp. Both premolars have distal and mesial roots that are compressed mesiodistally and incompletely separated by a deep lingual groove. The tall anterior corpus has a mandibular incisure and a labially positioned symphyseal tuber. The anterior symphyseal surface makes an angle of 70° with the alveolar margin. Lingually, an obliquely oriented post-incisive plane precedes a roughened genioglossus attachment, but there are no separate superior and inferior transverse tori. The tall but relatively narrow corpus has a single mental foramen below P_4 , and a strongly marked marginal torus with striae platysmatica that join distinct tubercles at C/P_3 and M_1/M_2 . The broad, tall ramus rises perpendicular to the occlusal plane and its root at M_2 defines a wide extramolar sulcus (10 mm). The notch between the condyle and the taller coronoid process is shallow; the latter markedly projects anteriorly (above M_2/M_3). The gonial angle is 100° ; this region is everted and extends below the corpus base causing a 'notched' appearance in lateral view. Muscle attachments are distinct: the masseter formed a thick everted flange infero-posteriorly and a clear fossa for its deep head; the marked medial pterygoid insertions extend to the mylohyoid line, which is partially bridged on the left. The mandibular foramen sits high above the occlusal plane.

KNM-ER 62003, with an estimated geological age of 1.90 to 1.95 Myr old, is a mandible fragment from just left of the midline to the distal right M_1 (Fig. 2; metrics in Supplementary Note 7). The midline inferior border is missing. The roots of the left I_1 to right M_1 are present, and the right P_3 to M_1 retain parts of their crown. Complete roots, thick cortical bone and substantial occlusal wear indicate that it is an adult or late sub-adult. KNM-ER 62003 shares the following features with KNM-ER 60000: a flat anterior arcade with short and narrow incisor roots, small premolars, similar symphysis orientation (the angle of the anterior surface is approximately 70°), an oblique post-incisive plane, and similar mental foramen and marginal torus positions. The marginal torus is less strongly developed in KNM-ER 62003 than in KNM-ER 60000, and a well-developed genial fossa delineates a superior transverse torus.

We attribute these mandibles to *Homo* rather than *Paranthropus* based on their small molars and premolars, and tall but mediolaterally narrow corpora. Despite similarities in P_3 – M_2 size, KNM-ER 60000 differs from *H. erectus* as it has a combination of smaller anterior teeth, a larger M_3 , a broader extramolar sulcus and, compared with the few known relatively complete adult and late juvenile mandibles (KNM-ER 992, KNM-WT 15000, D2735 and D2600), larger bigonial and bicondylar breadths for dental arcade size.

The discovery of a fossil (KNM-ER 62000) that is broadly contemporary with and very similar in shape to KNM-ER 1470, makes it likely that the two specimens do indeed represent a distinct taxon, separate from morphologically more generalized craniofacial specimens, such as KNM-ER 1813 and OH 65 (Supplementary Note 5). Importantly, although KNM-ER 1470 and KNM-ER 62000 span the full known size-range of early *Homo*, they share the same distinctive facial morphology, suggesting that allometry or sexual dimorphism^{5,6,8} are unlikely to be factors that underlie the differences in early *Homo* facial architecture.

The palate and teeth of KNM-ER 62000 make it possible to infer aspects of the mandible that occluded with the distinctive upper dental arcade of this specimen and KNM-ER 1470 (Fig. 3; Supplementary Note 10). KNM-ER 1802 is the mandible invariably grouped with KNM-ER 1470 (refs 3, 10, 12, 14–16), and is key to making associations with other specimens^{7,22}. However, when reconstructed it seems to be an unlikely match for KNM-ER 1470 and KNM-ER 62000, as it has a long dental arcade and an anteriorly arched incisor row. Instead, the new specimens

Figure 4 | A composite section of the Karari Ridge strata. The stratigraphic context of the three hominin fossils are shown (detailed individual sections are provided in Supplementary Note 13). The thickness of strata is given in metres.

KNM-ER 60000 and KNM-ER 62003, as well as mandibles KNM-ER 1482 and KNM-ER 1801, are a better match for KNM-ER 1470 and KNM-ER 62000 because of their shorter arcade, including a short pre-molar row, and a short, non-projecting incisor row (Supplementary Notes 10 and 11). Moreover, the pronounced anterior projection of the ramus of KNM-ER 60000 can be comfortably accommodated by the anteriorly placed zygomatic processes seen in the two crania. The species name of KNM-ER 1470, and of its associated specimens, depends on whether OH 7, the type specimen of *H. habilis*, is part of this group (see, for example, refs 15, 16, 19 for opposing views). If OH 7 is not part of this group, as suggested here by mandibular tooth shape (Supplementary Note 9) and perhaps by a photographic reconstruction of the mandible⁶, the name *H. rudolfensis* is available²³. However, we urge caution with respect to decisions on nomenclature until OH 7, which has extensive and complex taphonomic distortion, is studied further and its association with other fossils can be evaluated better.

KNM-ER 1470 provides the key evidence in studies recognizing two species of eastern African early *Homo*^{3,10–20}. Specimens have been grouped with KNM-ER 1470 mostly on the basis of a relatively large cranial size or inferred masticatory specialization^{3,12,15,16,19,20}, because it has no tooth crowns preserved, and because before the discovery of KNM-ER 62000 no other fossil shared its distinctive facial morphology. The new evidence presented here is not only consistent with two distinct craniofacial morphs but also with two mandibular morphs. One of these, best represented by KNM-ER 60000, seems to be associated with the species to which KNM-ER 1470 and KNM-ER 62000 belong. The three new specimens will greatly aid the reassessment of the systematics and early radiation of the genus *Homo*.

METHODS SUMMARY

Details of the new fossils are given in Supplementary Note 12. Their stratigraphic context within the Koobi Fora Formation²⁴ is well constrained (Fig. 4; Supplementary Notes 13), and their age derives from two temporal markers, the KBS Tuff and the Olduvai Subchron. The KBS Tuff has been isotopically dated to 1.87 Myr ago (ref. 25) and it outcrops within each of the stratigraphic sequences discussed here (Fig. 4). The base of the Olduvai Subchron, dated to 1.95 Myr ago (ref. 26), has recently been documented just to the north²⁷ and south²⁸ of these sections. KNM-ER 62003 was recovered from the lowermost upper Burgi Member in Area 130, just above a prominent intra-formational conglomerate (marker bed g), 14 m below the KBS Tuff, and very close to the projected level of the base of the Olduvai Subchron. Estimates of its age, based on stratigraphic scaling, give an age of 1.90 to 1.95 Myr old. KNM-ER 62000 was discovered in Area 131, in a comparable stratigraphic position 17 m below the KBS Tuff, deriving from a channel sand body that locally erodes through the same marker bed g. Its estimated age is 1.91 to 1.95 Myr old. KNM-ER 60000 was recovered in Area 105 from a broken sandstone block that derives from just above the KBS Tuff, and is therefore younger than 1.87 Myr old. On the basis of the local sections in Area 105, it is difficult to put a minimum age on the fossil, but magnetic polarity stratigraphy for that section²⁹ indicates that the fossil must lie within the Olduvai Subchron, and therefore cannot be younger than 1.78 Myr old²⁶. Finally, KNM-ER 1470, recovered from Area 131 and derived from 36 m below the KBS Tuff, now has an estimated age of 2.03 Myr old³⁰.

Received 25 March; accepted 12 June 2012.

1. Leakey, R. E. F. Evidence of an advanced Plio-Pleistocene hominid from East Rudolf, Kenya. *Nature* **242**, 447–450 (1973).
2. Gibbard, P. L. et al. Formal ratification of the quaternary system/period and the Pleistocene series/epoch with a base at 2.58 Ma. *J. Quaternary Sci.* **25**, 96–102 (2010).
3. Wood, B. *Koobi Fora Research Project Vol. 4: Hominid Cranial Remains* (Clarendon, 1991).
4. Howell, F. C. in *Evolution of African Mammals* (eds Maglio, V. J. & Cooke, H. B. S.) 154–248 (Harvard Univ. Press, 1978).
5. Johanson, D. C. et al. New partial skeleton of *Homo habilis* from Olduvai Gorge. *Nature* **327**, 205–209 (1987).
6. Tobias, P. V. *Olduvai Gorge Volume 4: The Skulls and Endocasts of Homo habilis*. (Cambridge Univ. Press, 1991).
7. Suwa, G., White, T. D. & Howell, F. C. Mandibular post-canine dentition from the Shungura Formation, Ethiopia: crown morphology, taxonomic allocation, and Plio-Pleistocene hominid evolution. *Am. J. Phys. Anthropol.* **101**, 247–282 (1996).
8. Miller, J. M. Craniofacial variation in *Homo habilis*: an analysis of the evidence for multiple species. *Am. J. Phys. Anthropol.* **112**, 103–128 (2000).

9. Suwa, G. et al. Early Pleistocene *Homo erectus* fossils from Konso, southern Ethiopia. *Anthropol. Sci.* **115**, 133–151 (2007).
10. Leakey, R. E. F. Further evidence of Lower Pleistocene hominids from East Rudolf, North Kenya, 1973. *Nature* **248**, 653–656 (1974).
11. Wood, B. A. in *Ancestors: The Hard Evidence* (ed. Delson, E.) 206–214 (Alan R. Liss, 1985).
12. Stringer, C. B. in *Major Topics in Primate and Human Evolution* (eds Wood, B., Martin, L. & Andrews, P.) 266–294 (Cambridge Univ. Press, 1986).
13. Lieberman, D. E., Pilbeam, D. R. & Wood, B. A. A probabilistic approach to the problem of sexual dimorphism in *Homo habilis*: A comparison of KNM-ER 1470 and KNM-ER 1813. *J. Hum. Evol.* **17**, 503–511 (1988).
14. Groves, C. P. *A Theory of Human and Primate Evolution* (Clarendon, 1989).
15. Wood, B. Origin and evolution of the genus *Homo*. *Nature* **355**, 783–790 (1992).
16. Rightmire, G. P. Variation among early *Homo* crania from Olduvai Gorge and the Koobi Fora region. *Am. J. Phys. Anthropol.* **90**, 1–33 (1993).
17. Kramar, A., Donnelly, S. M., Kidder, J. H., Ousley, S. D. & Olah, S. M. Craniometric variation in large-bodied hominids: testing the single-species hypothesis for *Homo habilis*. *J. Hum. Evol.* **29**, 443–462 (1995).
18. Kimbel, W. H., Johanson, D. C. & Rak, Y. Systematic assessment of a maxilla of *Homo* from Hadar, Ethiopia. *Am. J. Phys. Anthropol.* **103**, 235–262 (1997).
19. Blumenshine, R. J. et al. Late Pliocene *Homo* and hominid land use from western Olduvai Gorge, Tanzania. *Science* **299**, 1217–1221 (2003).
20. Clarke, R. J. A *Homo habilis* maxilla and other newly-discovered hominid fossils from Olduvai Gorge, Tanzania. *J. Hum. Evol.* <http://dx.doi.org/10.1016/j.jhevol.2011.11.007> (4 May 2012).
21. Dean, M. C. & Smith, B. H. in *The First Humans: Origin of the Genus Homo* (eds Grine, F. E., Fleagle, J. G. & Leakey, R. E.) 101–120 (Springer, 2009).
22. Bromage, T. G., Schrenk, F. & Zonneveld, F. W. Paleoanthropology of the Malawi Rift: An early hominid mandible from the Chiwondo Beds, northern Malawi. *J. Hum. Evol.* **28**, 71–108 (1995).
23. Wood, B. '*Homo rudolfensis*' Alexeev, 1986 – fact or phantom? *J. Hum. Evol.* **36**, 115–118 (1999).
24. Brown, F. H. & Feibel, C. S. Revision of lithostratigraphic nomenclature in the Koobi Fora region, Kenya. *J. Geol. Soc. Lond.* **143**, 297–310 (1986).
25. McDougall, I. & Brown, F. H. Precise ⁴⁰Ar/³⁹Ar geochronology for the upper Koobi Fora Formation, Turkana Basin, northern Kenya. *J. Geol. Soc. Lond.* **163**, 205–220 (2006).
26. Lourens, L., Hilgen, F., Shackleton, N. J., Laskar, J. & Wilson, D. in *A Geologic Time Scale* (eds Gradstein, F., Ogg, J. & Smith, A.) 409–440 (Cambridge Univ. Press, 2004).
27. Braun, D. R. et al. Early hominid diet included diverse terrestrial and aquatic animals 1.95 Ma in East Turkana, Kenya. *Proc. Natl Acad. Sci. USA* **107**, 10002–10007 (2010).
28. Joordens, J. C. A. et al. An astronomically-tuned climate framework for hominins in the Turkana Basin. *Earth Planet. Sci. Lett.* **307**, 1–8 (2011).
29. Brock, A., Isaac, G. & L.L. Paleomagnetic stratigraphy and chronology of hominid-bearing sediments east of Lake Rudolf, Kenya. *Nature* **247**, 344–348 (1974).
30. McDougall, I. et al. New single crystal ⁴⁰Ar/³⁹Ar ages improve time scale for deposition of the Omo Group, Omo-Turkana Basin, East Africa. *J. Geol. Soc. Lond.* **169**, 213–226 (2012).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Governments of Kenya and Tanzania for permission to carry out this research, the Kenya Wildlife Service for permission to work in the Sibiloi National Park, the National Museums of Kenya and the National Museum of Tanzania for access to specimens in their care, and the Turkana Basin Institute for support. The National Geographic Society, the Leakey Foundation and the Max Planck Society funded fieldwork or laboratory studies. Many people helped us with this research, including N. Adamali, R. Blumenshine, C. Boesch, F. Brown, P. Gunz, J. J. Hublin, W. Kimbel, K. Kupczik, R. Leakey, C. Lepre, D. Lieberman, P. Mserwa, R. Odoyo, R. Quinn, P. Rightmire, L. Schroeder, U. Schwarz, M. Skinner, H. Temming, A. Winzer and B. Wood. Curatorial assistance was given by A. Kweka, F. Manthi, E. Mbua, M. Muungu and J. Thiringi. KNM-ER 60000 was discovered by C. Nyete, KNM-ER 62000 by D. Elgite and KNM-ER 62003 by R. Moru. We particularly thank the Koobi Fora Research Project field crew: A. Aike, S. Aila, D. Elgite, M. Kirinya, D. Gidole, O. Kyalo, A. Longaye, A. Lawri, E. Linga, J. Lonyericho, S. Lomeiku, D. Muema, A. Moru, R. Moru, S. Muge, C. Nyete, L. Nzuve, H. Sale and A. Sharamo whose fieldwork led to the discovery of these specimens, and camp managers J. Mutuku and T. Ngundo. H. Churcher, J. Coreth, A. Hammond, J. LaCarrubba, F. Kirera, C. Lepre, M. Noback, R. Quinn, M. Skinner, I. Wallace and S. Wright participated in one or more of the 2007, 2008 or 2009 field expeditions when these specimens were discovered. We are grateful to F. and J. Pinto, W. Phillips, M. Hettwer, P. Sylvester, H. Buchi, N. Seligman, E. von Simpson, J. Doerr and B. and J. Chelberg for their financial support of this fieldwork.

Author Contributions M.G.L. and L.N.L. directed the field research, in which C.S.F. and F.S. participated. C.K. and F.S. prepared the hominid fossils, F.S. and M.C.D. made the virtual reconstructions, and C.S.F. studied the geological context. M.G.L., F.S., M.C.D., S.C.A. and L.N.L. made comparative observations and carried out analyses. F.S. took the lead in writing the paper, and S.C.A., M.C.D. and C.S.F. contributed.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.G.L. (meaveleakey@gmail.com) or F.S. (fspoor@eva.mpg.de).

A transitional snake from the Late Cretaceous period of North America

Nicholas R. Longrich¹, Bhart-Anjan S. Bhullar² & Jacques A. Gauthier¹

Snakes are the most diverse group of lizards¹, but their origins and early evolution remain poorly understood owing to a lack of transitional forms. Several major issues remain outstanding, such as whether snakes originated in a marine^{2–4} or terrestrial^{5,6} environment and how their unique feeding mechanism evolved^{1,7,8}. The Cretaceous *Coniophis precedens* was among the first Mesozoic snakes discovered⁹, but until now only an isolated vertebra has been described^{9,10} and it has therefore been overlooked in discussions of snake evolution. Here we report on previously undescribed material¹¹ from this ancient snake, including the maxilla, dentary and additional vertebrae. *Coniophis* is not an anilioid as previously thought¹¹; a revised phylogenetic analysis of Ophidia shows that it instead represents the most primitive known snake. Accordingly, its morphology and ecology are critical to understanding snake evolution. *Coniophis* occurs in a continental floodplain environment, consistent with a terrestrial rather than a marine origin; furthermore, its small size and reduced neural spines indicate fossorial habits, suggesting that snakes evolved from burrowing lizards. The skull is intermediate between that of lizards and snakes. Hooked teeth and an intramandibular joint indicate that *Coniophis* fed on relatively large, soft-bodied prey. However, the maxilla is firmly united with the skull, indicating an akinetic rostrum. *Coniophis* therefore represents a transitional snake, combining a snake-like body and a lizard-like head. Subsequent to the evolution of a serpentine body and carnivory, snakes evolved a highly specialized, kinetic skull, which was followed by a major adaptive radiation in the Early Cretaceous period. This pattern suggests that the kinetic skull was a key innovation that permitted the diversification of snakes.

By almost any standard, snakes are among the most successful vertebrate groups. Extant Serpentes are represented by almost 3,000 species exploiting a remarkable range of habitats and prey¹. However, the origins of this radiation remain poorly understood. In particular, there is a longstanding controversy over whether the elongate body and reduced limbs of snakes evolved in a terrestrial setting^{5,6}, perhaps as an adaptation for burrowing, or in a marine environment^{2–4}, as an adaptation for swimming. The stem snake *Najash* is terrestrial⁵, arguing against a marine origin, yet the animal's large size (~2 m in length¹²) seems inconsistent with the hypothesis that snakes evolved from burrowing forms. Another problem concerns the evolution of the specialized feeding apparatus of snakes⁸. Extant snakes are primarily carnivores that feed on relatively large prey, a strategy facilitated by highly flexible jaws^{1,7,8,13}. However, fossils provide little information on the transition from the primitive, relatively inflexible skulls of other lizards to the derived, kinetic system of snakes⁷, or when this transition occurred in snake evolution.

Answers to these questions can only come from fossils, but the fossil record of snakes is highly incomplete. Early snakes are known mainly from vertebrae, which provide little information about snake evolution. *Coniophis precedens*, from the Maastrichtian stage of North America^{9–11}, is one such species. Described by Marsh in 1892 from a

single vertebra⁹, *Coniophis* has traditionally been identified as a primitive alethinophidian¹⁴, perhaps a pipe snake¹¹ (Aniliidae), although it has also been suggested that *Coniophis* represents a stem snake¹⁵. Surprisingly, additional vertebrae and skull elements of *Coniophis* have been collected¹¹, but never described. Here we provide such a description and show that *Coniophis* is intermediate between snakes and other lizards.

All of the fossils that we describe here come from the same strata as the holotype, the upper Maastrichtian Lance Formation of eastern Wyoming. Vertebrae referred to *Coniophis* resemble the holotype in possessing rudimentary neural spines and deep, obovate paracotylar fossae. Cranial elements come from localities containing numerous *Coniophis* vertebrae¹¹. Referral is supported by occurrence, size, relative abundance (Supplementary Information, section 1) and the fact that elements occupy similar phylogenetic positions when analysed independently (Supplementary Information, section 2).



Figure 1 | Dentary of *Coniophis precedens*. UCMF (University of California Museum of Paleontology) 50000 in lateral (a), medial (b), ventral (c) and dorsal (d) view. mec, Meckelian canal; mf, mental foramen; sp, splenial facet; sr, subdental ridge; sur, surangular notch; th, thecae.

¹Department of Geology and Geophysics, Yale University, PO Box 208109, New Haven, Connecticut 06520-8109, USA. ²Department of Organismic and Evolutionary Biology, Biological Laboratories, 16 Divinity Avenue, Harvard University, Cambridge, Massachusetts 02138 USA.

Jaw elements of *Coniophis* have a mosaic of derived, snake-like characters and primitive, lizard-like characters. Teeth are snake-like in being tall, pointed and cylindrical with a sharp basal hooking of the crown. This tooth shape is unique to snakes among squamates. Distinct carinae are present, as in Varanoidea and Mosasauroidea. Tooth bases are expanded; some bear striae but, unlike in Varanoidea and Madtsoiidae¹⁶, grooves do not extend into the pulp cavity. Each tooth has a basal nutrient foramen in the interdental position, as in other snakes and Anguimorpha. Tooth implantation is uniquely snake-like, with teeth attaching to shallow thecae separated by interdental ridges. However, unlike in alethinophidians, the thecae remain open lingually, and teeth lack extensive basal deposition of bone of attachment.

The dentary (Fig. 1) is snake-like in being long, slender and bowed. It bears an enlarged mental foramen, which represents a snake autapomorphy. Caudally, the dentary is deeply notched to receive the surangular. This feature is unique to snakes and is associated with an intramandibular joint permitting the jaw to flex to expand the gape^{8,17,18}. The subdental surface is deep anteriorly, as in snakes, but forms a thin lip posteriorly, which is a primitive feature not found in any snake, including the basal snakes *Najash*¹² and *Dinilysia*¹⁹. The angular process of the dentary does not wrap beneath the Meckelian fossa, a lizard-like feature otherwise seen only in *Najash*¹² among snakes. A facet along the back of the subdental ridge marks the splenial articulation.

The maxilla (Fig. 2) is low and medially curved as in other snakes, indicating a broad, U-shaped snout; a low facial process is present, as in *Dinilysia*¹⁹ and Madtsoiidae^{20,21}. The maxilla is lizard-like, but differs from all other snakes in that it overlaps the premaxilla, producing an immobile contact. *Coniophis* also differs from other snakes in retaining a vomerine process that is distally expanded to articulate with the palate. The preserved narial margin indicates a small, unretracted external naris. Overlapping portions of the maxilla indicate at least

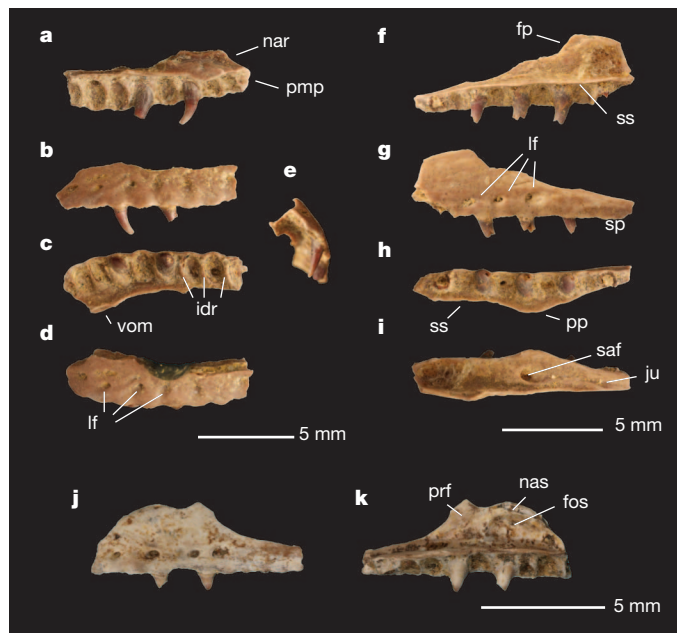


Figure 2 | Maxilla of *Coniophis precedens*. a–e, UCM 53935, anterior part of maxilla in medial (a), lateral (b), ventral (c), dorsal (d) and anterior (e) view. f–i, UCM 49999, posterior maxilla in medial (f), lateral (g), ventral (h) and dorsal (i) view. j, k, AMNH (American Museum of Natural History) 22413, posterior maxilla in lateral (j) and medial (k) view. fos, fossa for nasal capsule; fp, facial process; idr, interdental ridge; ju, jugal articulation; lf, labial foramina; nar, narial margin; nas, nasal contact; pp, palatine process; pmp, premaxillary process; prf, prefrontal facet; saf, superior alveolar foramen; ss, supradental shelf; vom, vomerine process.

nine labial foramina; other snakes have five at most. Dorsally, the facial process bears a groove for the nasal and a lappet for the prefrontal, indicating immobile contacts with these bones, in contrast to all other snakes. Medially, the nasolacrimal fossa is lost, but *Coniophis* retains a well-developed, lizard-like fossa for the nasal capsule. A distinct supradental shelf separates the facial process from the teeth, whereas this shelf is lost anteriorly in Serpentes. The superior alveolar foramen is lizard-like in opening dorsally, rather than medially as in Alethinophidia. The suborbital process bears a V-shaped notch for the jugal, unlike any known snake.

Coniophis vertebrae (Fig. 3) are common enough in the Lance Formation to allow characterization of the morphology along the snake's axial column. Snake-like zygosphenes and zygantra are developed throughout; the neural canal is snake-like in having an inverted trefoil outline, with longitudinal canals ventrolaterally. Condyles are taller than they are wide anteriorly and are depressed posteriorly, as in Scolecophidia, and lack the groove around the condyle seen in alethinophidians. Anterior vertebrae bear stubby hypapophyses, as in *Dinilysia*²² and *Najash*¹², mid-trunk vertebrae bear a sharp keel and posterior vertebrae have a rounded ventral surface. Synapophyses are single-headed, as in lizards and Scolecophidia; they bear diagnostic deep, obovate paracotylar fossae. In middle and anterior vertebrae, synapophyses lie level with the ventral margin of the centrum, as in snakes; but in posterior vertebrae they lie above the centrum, as in lizards but no other snake. Prezygapophyseal processes are absent.

The neural arch is depressed, with a straight to weakly concave caudal margin, which is a primitive feature. The neural spine is reduced to a weak ridge terminating posteriorly in a tuberosity. On either side of the neural spine is a low mound corresponding to the dorsolateral ridges in *Najash*¹², *Dinilysia*²² and Madtsoiidae¹⁸. A vertebra from the tail bears lymphapophyses for support of lymph hearts, as in other long-bodied squamates²³, indicating that *Coniophis* had a serpentine body.

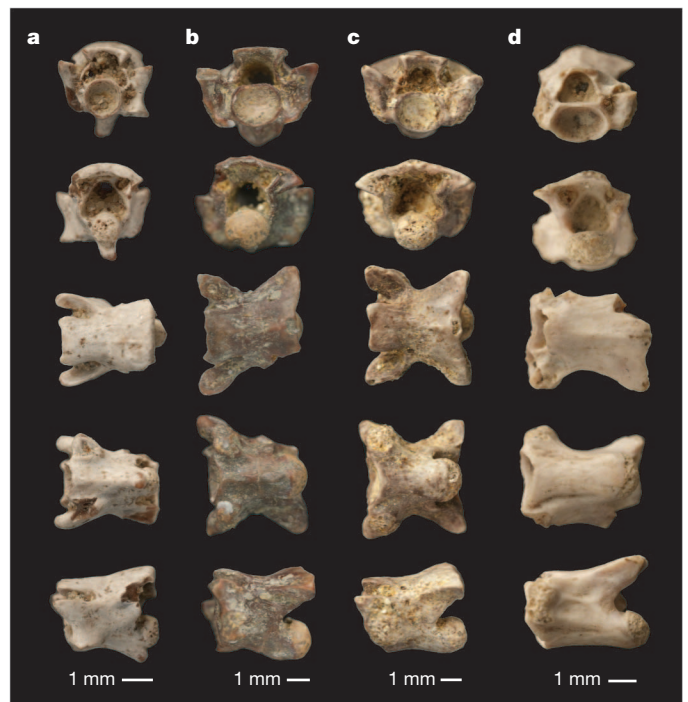


Figure 3 | Vertebrae of *Coniophis precedens*. a, AMNH 26999, vertebra from cervical region. b, USNM (United States National Museum) 2143, holotype anterior trunk vertebra. c, AMNH 26833, middle trunk vertebra. d, YPM-PU (Yale Peabody Museum, Princeton collection) 16845, posterior trunk vertebra. Shown (from top to bottom) in anterior, posterior, dorsal, ventral and lateral views.

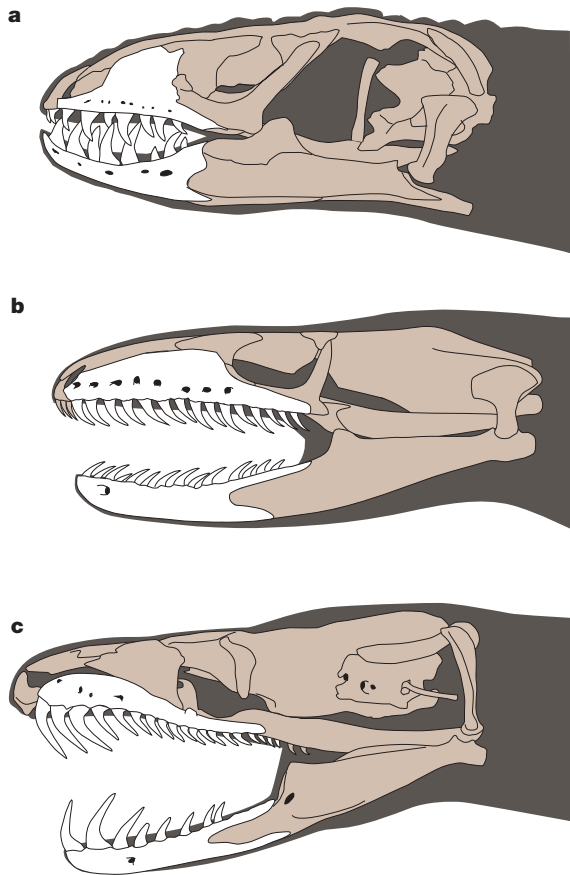


Figure 4 | Skull in lizards, *Coniophis* and modern snakes. a, *Heloderma* (Varanoidea). b, Reconstruction of *Coniophis*, with missing elements after Anguillidae and basal snakes *Najash* and *Dinilysia*. c, *Epicrates* (Macrostromata).

Phylogenetic analysis places *Coniophis precedens* as the sister to all known snakes. Similar topologies emerge whether we code *Coniophis* elements together or separately, add *Amphisbaenia* to the outgroup or

use another matrix¹⁸ entirely (Supplementary Information, section 2). Numerous synapomorphies, including hooked teeth, interdental ridges, an intramandibular joint, an anteriorly deep subdental ridge, a reduced maxillary facial process, loss of the lacrimal fossa and zygosphene–zygtrum joints (Supplementary Information, section 2), place *Coniophis* with Ophidia. However, the new skull material shows that *Coniophis* retains several strikingly primitive features absent from crown Serpentes, including a supradental shelf, a vomerine process and tight sutural contacts between the maxilla and the premaxilla, nasals, prefrontal and jugal. Six steps are required to place *Coniophis* in the crown, and 16 steps are needed to place *Coniophis* with anilioids.

As the most basal known snake, *Coniophis precedens* is important to reconstructing the origins of the distinctive ophidian body plan and the ecology of ancestral snakes. The proposed marine origin of snakes^{2–4} suggests that basal snakes will occur in marine deposits. However, *Coniophis* occurs in continental floodplain sediments^{11,24}, as do other basal snakes^{3,25}. *Coniophis* also lacks adaptations for aquatic locomotion such as pachyostosis², tall neural arches or ventrally placed synapophyses²⁴. Instead, vertebrae have depressed neural arches with reduced neural spines, a feature shared with fossorial snakes⁵ such as *Scolecophidia* and *Anilioidea*, and other burrowing squamates such as *Amphisbaenia*²⁶. This, together with the relatively small size of *Coniophis* (on the basis of comparisons of centrum length with anilioids, we estimate the snout–vent length to be ~700 mm), implies a semifossorial or fossorial lifestyle. *Coniophis* therefore supports the hypothesis that the elongate body and reduced limbs of snakes evolved as burrowing adaptations.

Coniophis also offers insight into the evolution of snake feeding. The teeth of *Coniophis* are suited to piercing and holding soft-bodied prey; moreover, the intramandibular joint allows the jaws to flex to increase gape size^{8,17,18}. Although it has been proposed that early snakes fed on invertebrates^{1,8}, taken together the tooth morphology, intramandibular joint and size of *Coniophis* suggest that it preyed upon small vertebrates. However, the maxilla lacks modifications permitting kinesis, and instead remained tightly bound to the premaxilla, vomer, nasal, prefrontal and jugal. *Coniophis* thus represents a functional chimaera, combining a snake-like body with a lizard-like head (Fig. 4).

The picture that emerges for *Coniophis* is of a small, fossorial carnivore that preyed upon small vertebrates. Although vertebral

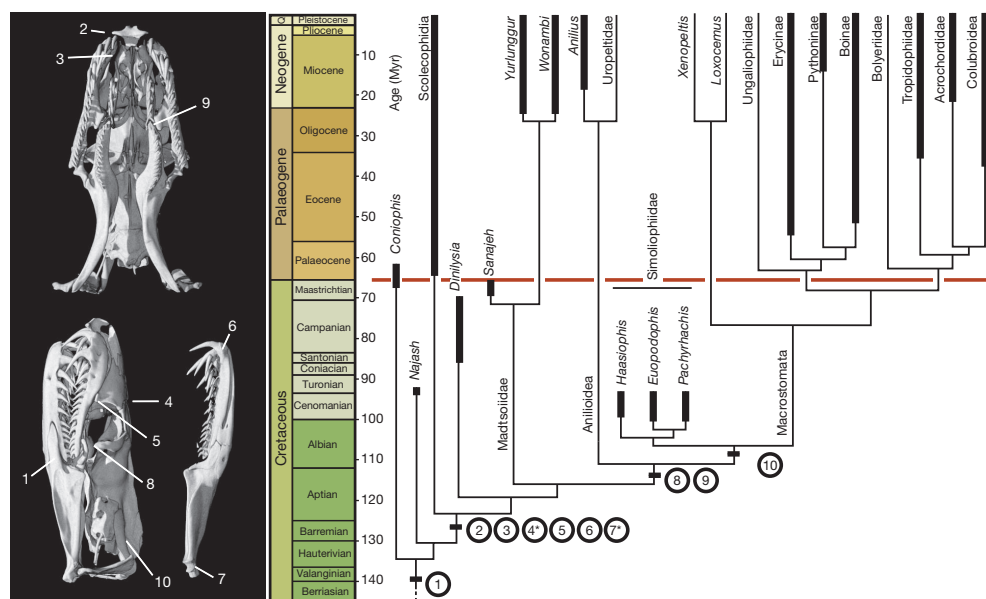


Figure 5 | Phylogeny of Ophidia, showing relationships of *Coniophis* and evolution of cranial kinesis. Adaptations permitting cranial kinesis (shown at left in the boa *Epicrates*) cluster near the base of the snake tree, but following the divergence of *Coniophis*. *Coniophis* exhibits an intramandibular joint (1). Serpentes exhibits a maxilla–premaxilla joint (2), loss of maxilla–vomer contact

(3), a nasofrontal joint (4), a maxilla–prefrontal joint (5), a mobile dentary symphysis (6) and an articular saddle joint (7). Alethinophidia is characterized by a reduced postorbital bar (8) and a palatine–pterygoid hinge (9). Macrostromata is characterized by a hinged supratemporal (10). Characters 4* and 7* are unknown for either *Coniophis* or *Najash*. Stratigraphic data from ref. 18.

morphology suggests a degree of fossoriality, a vertebrate diet implies foraging above ground. Many basal Alethinophidia have a similar lifestyle¹, and this ecology is probably ancestral for snakes. Later, snakes evolved a suite of adaptations that facilitated kinesis in the jaws and skull (Fig. 5), allowing a wider range of prey to be swallowed⁸. Subsequently, snakes underwent an adaptive radiation in the Early Cretaceous (Fig. 5). This radiation was well under way by the Cenomanian stage²⁷, by which point snakes were represented by fossorial, *Coniophis*-like forms^{28,29}, freshwater Nigerophiidae²⁸, marine Simoliophiidae²⁷ and the large-bodied, carnivorous Madtsoiidae²⁸. According to the phylogeny presented here, the aquatic Simoliophiidae, the giant Madtsoiidae and the insectivorous Scolecophidia do not provide insight into stem snake ecology; instead they are specialized offshoots of this initial diversification. *Coniophis* therefore shows how the stepwise acquisition of key innovations culminated in a major adaptive radiation. The genesis of the Serpentes began with the evolution of a novel means of locomotion, followed by adaptations facilitating the ingestion of ever larger prey, thereby enabling snakes to exploit a wider range of ecological niches.

METHODS SUMMARY

To determine the phylogenetic position of *Coniophis*, we revised a recent phylogenetic analysis of snake morphology¹⁹, adding 61 characters to resolve the position of *Coniophis* and the overall tree topology. We follow that study in including Uropeltinae, *Anomochilus* and *Cylindrophis* in Uropeltidae¹⁹, although the monophyly of this assemblage is debated. The resulting matrix (Supplementary Information, section 4) includes 23 taxa and 215 characters (Supplementary Information, sections 2 and 3). Data were analysed using the branch-and-bound search algorithm of PAUP* 4.0 b10 (<http://paup.csit.fsu.edu/>). Three additional analyses were performed: an analysis examining the maxillae, dentary and vertebrae separately, along with two unnamed Maastrichtian snakes; an analysis including the amphisbaenian *Rhineura floridana* in the outgroup; and an analysis using another matrix¹⁸. All results support placement of *Coniophis* as a basal snake (Supplementary Information, section 2).

Received 28 March; accepted 11 May 2012.

Published online 25 July 2012.

- Greene, H. W. *Snakes: the Evolution of Mystery in Nature* (Univ. California Press, 1997).
- Caldwell, M. W. & Lee, M. S. Y. A snake with legs from the marine Cretaceous of the Middle East. *Nature* **386**, 705–709 (1997).
- Cope, E. D. On the reptilian orders Pythonomorpha and Streptosauria. *Proc. Bost. Soc. Nat. Hist.* **12**, 250–267 (1869).
- Lee, M. S. Y. Molecular evidence and marine snake origins. *Biol. Lett.* **1**, 227–230 (2005).
- Apesteguía, S. & Zaher, H. A Cretaceous terrestrial snake with robust hindlimbs and a sacrum. *Nature* **440**, 1037–1040 (2006).
- Vidal, N. & Hedges, S. B. Molecular evidence for a terrestrial origin of snakes. *Proc. R. Soc. Lond. B* **271**, S226–S229 (2004).
- Lee, M. S. Y., Bell, G. L. & Caldwell, M. W. The origin of snake feeding. *Nature* **400**, 655–659 (1999).
- Cundall, D. & Greene, H. W. in *Feeding: Form, Function, and Evolution in Terrestrial Vertebrates* (ed. Schwenk, K.) 293–333 (Academic, 2000).
- Marsh, O. C. Notice of new reptiles from the Laramie Formation. *Am. J. Sci.* **43**, 449–453 (1892).
- Gilmore, C. W. Fossil snakes of North America. *Geol. Soc. Am. Bull.* **9**, 1–96 (1938).

- Estes, R. Fossil Vertebrates from the Late Cretaceous Lance Formation, Eastern Wyoming. *Univ. Calif. Publ. Geol. Sci.* **49**, 140–141 (1964).
- Zaher, H., Apesteguía, S. & Scanferla, C. A. The anatomy of the Upper Cretaceous snake *Najash rionegrina* Apesteguía & Zaher, 2006, and the evolution of limblessness in snakes. *Zool. J. Linn. Soc.* **156**, 801–826 (2009).
- Cundall, D. & Irish, F. in *Biology of the Reptilia* Vol. 20 (eds Gans, C., Gaunt, A. S. & Adler, K.) 349–692 (Society for the Study of Amphibians and Reptiles, 2008).
- Hoffstetter, R. in *Traité de Paléontologie* (ed. Piveteau, J.) 606–662 (Maison et Cie, 1955).
- Rage, J.-C. & Augé, M. Squamate reptiles from the middle Eocene of Lissieu (France). *Geobios* **43**, 253–268 (2010).
- Scanlon, J. D. & Lee, M. S. Y. Varanoid-like dentition in primitive snakes (Madtsoiidae). *J. Herpetol.* **36**, 100–106 (2002).
- Cundall, D. Feeding behavior in *Cylindrophis* and its bearing on the evolution of alethinophidian snakes. *J. Zool.* **237**, 353–376 (1995).
- Wilson, J. A., Mohabey, D. M., Peters, S. E. & Head, J. J. Predation upon hatchling dinosaurs by a new snake from the Late Cretaceous of India. *PLoS Biol.* **8**, e1000322 (2010).
- Zaher, H. & Scanferla, C. A. The skull of the Upper Cretaceous snake *Dinilysia patagonica* Smith-Woodward 1901, and its phylogenetic position revisited. *Zool. J. Linn. Soc.* **164**, 194–238 (2012).
- Scanlon, J. D. & Lee, M. S. Y. The Pleistocene serpent *Wonambi* and the early evolution of snakes. *Nature* **403**, 416–420 (2000).
- Scanlon, J. D. Skull of the large non-macrostromatan snake *Yurlunggur* from the Australian Oligocene. *Nature* **439**, 839–842 (2006).
- Rage, J.-C. & Albino, A. M. *Dinilysia patagonica* (Reptilia, Serpentes): matériel vertébral additionnel du Crétacé supérieur d'Argentine. Étude complémentaire des vertèbres, variations intraspécifiques et intracolumnaires. *Neues Jahrb. Geol. Paläontol. Monatsh.* **1989**, 433–447 (1989).
- Hoffstetter, R. & Gasc, J.-P. Vertebrae and ribs of modern reptiles. in *Biology of the Reptilia* Vol. 1 (ed. Gans, C.) 201–310 (Academic, 1969).
- Prasad, G. V. R. & Rage, J.-C. Amphibians and squamates from the Maastrichtian of Naskal, India. *Cretac. Res.* **16**, 95–107 (1995).
- Caldwell, M. W. & Albino, A. M. Palaeoenvironment and palaeoecology of three Cretaceous snakes: *Pachyophis*, *Pachyrhachis*, and *Dinilysia*. *Acta Palaeontol. Pol.* **46**, 203–218 (2001).
- Kearney, M. Systematics of the Amphisbaenia (Lepidosauria: Squamata) based on morphological evidence from recent and fossil forms. *Herpetological Monogr.* **17**, 1–74 (2003).
- Rage, J.-C. & Escuil, F. The Cenomanian: stage of hindlimbed snakes. *Carnets Géol.* **2003**, 2003/01 (2003).
- Rage, J.-C. & Werner, C. Mid-Cretaceous (Cenomanian) snakes from Wadi Abu Hashim, Sudan: the earliest snake assemblage. *Palaeontol. Afr.* **35**, 85–110 (1999).
- Gardner, J. D. & Cifelli, R. L. A primitive snake from the Cretaceous of Utah. *Spec. Pap. Palaeontol.* **60**, 87–100 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the curators and staff of the American Museum of Natural History, the University of California Museum of Paleontology and the Yale Peabody Museum of Natural History, and the staff of Digimorph at the University of Texas at Austin, for specimen access. We are especially grateful to J. Head for discussions about snake vertebrae, which prevented a major error in the assignment of material to *Coniophis*. This research was funded in part by US NSF grant DEB-0132227 to J.A.G. N.R.L. was funded by the Yale Institute for Biospheric Studies.

Author Contributions N.R.L. designed the research, identified specimens, collected data, performed the phylogenetic analysis and wrote the paper. B.-A.S.B. designed the research, collected data and wrote the paper. J.A.G. collected data, contributed data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to N.R.L. (nicholas.longrich@yale.edu).

Atmospheric CO₂ forces abrupt vegetation shifts locally, but not globally

Steven I. Higgins¹ & Simon Scheiter²

It is possible that anthropogenic climate change will drive the Earth system into a qualitatively different state¹. Although different types of uncertainty limit our capacity to assess this risk², Earth system scientists are particularly concerned about tipping elements, large-scale components of the Earth system that can be switched into qualitatively different states by small perturbations. Despite growing evidence that tipping elements exist in the climate system^{1,3}, whether large-scale vegetation systems can tip into alternative states is poorly understood⁴. Here we show that tropical grassland, savanna and forest ecosystems, areas large enough to have powerful impacts on the Earth system, are likely to shift to alternative states. Specifically, we show that increasing atmospheric CO₂ concentration will force transitions to vegetation states characterized by higher biomass and/or woody-plant dominance. The timing of these critical transitions varies as a result of between-site variance in the rate of temperature increase, as well as a dependence on stochastic variation in fire severity and rainfall. We further show that the locations of bistable vegetation zones (zones where alternative vegetation states can exist) will shift as climate changes. We conclude that even though large-scale directional regime shifts in terrestrial ecosystems are likely, asynchrony in the timing of these shifts may serve to dampen, but not nullify, the shock that these changes may represent to the Earth system.

Assessing the consequences of climate change for life on Earth is one of the major challenges facing science and society. There is growing evidence that tipping elements, large-scale components of the Earth system that can be switched into qualitatively different states by small perturbations, exist in the climate system^{1,3}. There is also evidence that ecosystems may be subject to analogous tipping phenomena. Such regime shifts have been observed locally⁵, but whether such shifts occur at regional scales remains speculative. The terrestrial ecosystem where regime shifts are most likely to have powerful repercussions on the Earth system is the vast savanna complex, a term we use to refer to grassland, savanna and forest ecosystems of tropical and subtropical regions. Regime shifts of relevance to the Earth system are likely in the savanna complex because of the large area it covers and because savannas are known to shift into alternative states such as forest or grassland^{6–10}. Indeed, the stability of savannas has long been questioned by biogeographers, who noticed that they do not seem to be in equilibrium with their climate in the same way that forests seem to be¹¹. The Amazon basin is prominent example of a region in the savanna complex where shifts to alternative vegetation states are predicted. Several studies have shown that decreased rainfall over the Amazon would induce forest dieback, a process that would be reinforced and accelerated by feedbacks with the climate system^{1,3,12,13}. Despite the fact that large proportions of the savanna complex seem to

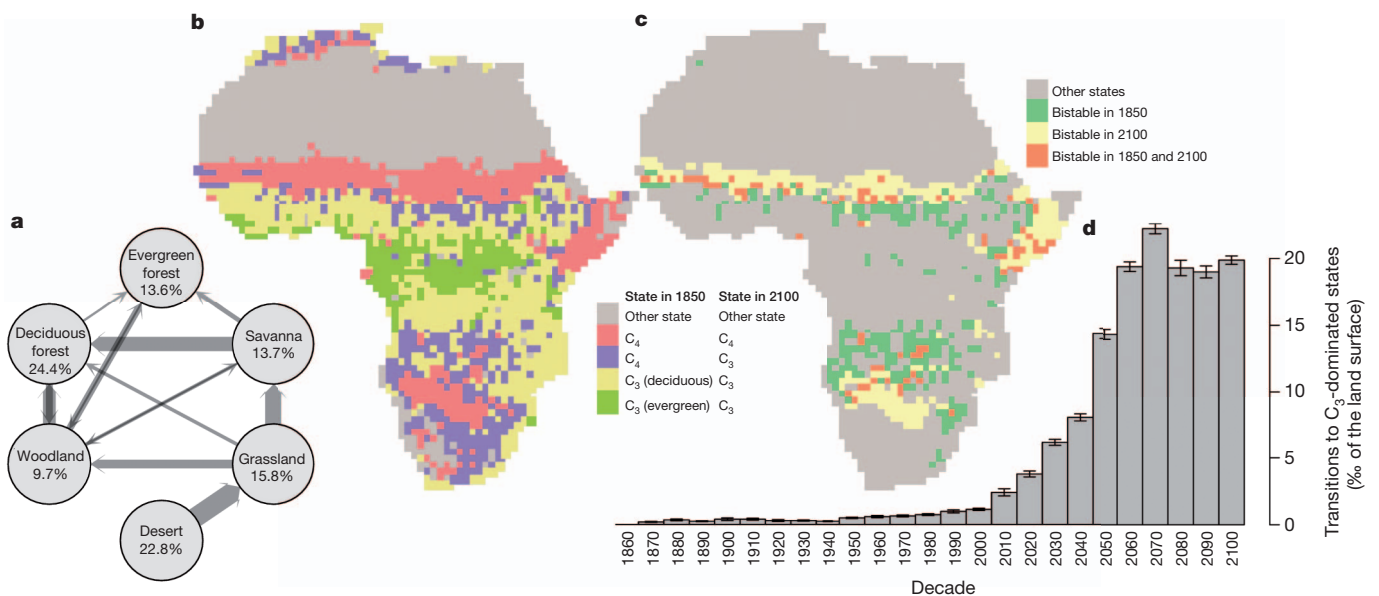


Figure 1 | Transitions between vegetation states projected for the period 1850–2100. **a**, The projected changes in vegetation states. The percentages indicate the proportion of the African surface that will be in different vegetation states in 2100. The width of the arrows shows the relative frequency of the vegetation transitions projected to occur between 1850 and 2100. **b**, The distribution of the most common state transitions between C₃-dominated

states (woodland, deciduous forest and evergreen forest) and C₄-dominated states (C₄ grassland and savanna). **c**, The distribution of bistable zones in 1850 and 2100. Bistable zones are savannas and C₄ grasslands that would undergo a transition to C₃-dominated states if fire were suppressed. **d**, Frequency distribution of the timing of transitions from C₄-dominated to C₃-dominated vegetation states. Error bars indicate s.e.m. for replicate simulations (n = 9).

¹Institut für Physische Geographie, Goethe Universität Frankfurt am Main, Altenhöferallee 1, 60438 Frankfurt am Main, Germany. ²Biodiversity and Climate Research Centre (BiK-F), Senckenberg Gesellschaft für Naturforschung, Senckenberganlage 25, 60325 Frankfurt am Main, Germany.

be predisposed to regime shifts^{9,10}, uncertainty in how climate and land use will change in these regions, coupled with uncertainty in the mechanisms driving change, hampers our ability to predict change¹⁴.

Addressing change in the savanna complex requires consideration of the physiological and demographic mechanisms that have been invoked to explain savanna dynamics. A mechanistic perspective needs to consider how the fundamentally different architectures of grasses and trees influence both their capacity to capture and use resources and their responses to fire¹⁵. A key physiological difference is that trees use the C₃ photosynthetic pathway, whereas most tropical grasses use the C₄ photosynthetic pathway. These differences yield robust predictions¹⁶: high temperatures and low CO₂ concentrations select for the C₄ photosynthetic pathway, whereas low temperatures and high CO₂ concentrations select for the C₃ photosynthetic pathway. According to leaf-level physiological models based on this mechanism¹⁶, a doubling of the CO₂ concentration from 350 to 700 p.p.m. and a 2–5 °C increase in temperature will favour C₃ over C₄ photosynthetic types. The implication is that, at the leaf level, the impacts of predicted CO₂ changes overwhelm the impacts of predicted temperature changes. In addition to carbon uptake, differences in carbon demand and carbon allocation need to be considered¹⁷. Trees in savannas have a higher carbon demand than grasses because, being trees, they require more carbon to deploy a unit of leaf area than grasses do and they need to allocate large amounts of carbon to ensure that they rapidly attain sizes that prevent serious injury in grass fires^{17,18}. In summary, the carbon budget of trees in savannas stands to benefit from increased atmospheric CO₂ concentration because of increased carbon uptake capacity fuelled by a high carbon demand. These carbon benefits setup a positive feedback loop: more tree biomass serves to suppress grass production and hence the severity of fires, which further increases the likelihood of the system tipping from a C₄-dominated state to a C₃-dominated state^{9,19}.

In this study we used a dynamic vegetation model²⁰ (see Supplementary Information) to assess vegetation shifts in grassland, savanna and forest formations that constitute the savanna complex. The model allows the assessment of shifts in the state of grassland, savanna and forest formations by explicitly considering how carbon assimilation may respond to atmospheric changes and how growth and resource competition interact with plant responses to injuries caused by fire (see Supplementary Information).

Our simulations of potential vegetation project marked shifts in vegetation state between 1850 and 2100 (Fig. 1a). There was a strong directionality in the shifts: deserts were replaced by grasslands, grasslands by savannas and woodlands, and savannas by forests (Fig. 1a). Collectively, the area occupied by C₃-dominated (woodland, deciduous forest and evergreen forest) states increased from 31% to 47%. The area covered by savannas decreased from 23% to 14%, grasslands decreased from 18% to 16% and deserts from 28% to 23%. There was a clear geographical pattern to where these transitions occurred (Fig. 1b). In areas associated with and bounding on what are referred to as hot semi-arid climates in the Köppen–Geiger climate classification system²¹, the model projects that C₄-dominated states (savanna and C₄ grassland) will undergo a transition to C₃-dominated states (woodland, deciduous forest and evergreen forest). Further, the positions of bistable areas (areas that could be C₃-dominated in the absence of fire, but are C₄-dominated if fires occur) are projected to shift in location (Fig. 1c). Only a small overlap in the past and future distributions of bistable areas is projected. Moreover, most sites that are projected to remain C₄-dominated (Fig. 1b, red pixels) are projected to switch to bistable states in 2100 (Fig. 1c, yellow pixels).

These marked shifts to woodland and forest have, according to the model, been occurring at low rates since 1850, but the rate of transition accelerates from 1.9‰ of the African land surface between 1990 and 2010 to 4.5‰ between 2010 and 2030, and to 11.2‰ between 2030 and 2050 (Fig. 1d). Although stochastic variation in the intensity of rainfall and fire events ensured that the confidence interval of a site's year of

transition spanned 48 years, the mean year when a site underwent a transition to a C₃ state was largely determined by the rate of temperature change at that site ($F_{2,215} = 218.5$, $P < 0.001$, $R^2 = 0.67$). That is, those sites currently dominated by C₄ plants in which temperatures are projected to increase more rapidly shifted to C₃ dominance later. This is because the relatively rapid increase in temperature allows the C₄ plants to remain competitive for longer in the face of rising atmospheric CO₂ concentrations, which favour C₃ plants¹⁶.

To understand how moisture and atmospheric CO₂ interact to influence these changes (Fig. 1) we plot how the probability of vegetation being in a C₄-dominated state changes with CO₂ concentration and rainfall (Fig. 2a). This plot reveals that at high CO₂ concentrations, C₄-dominated states become more likely at sites with low rainfall (less than 250 mm), which explains why the model predicts that grasslands replace deserts (Fig. 1a). This interpretation is consistent with the

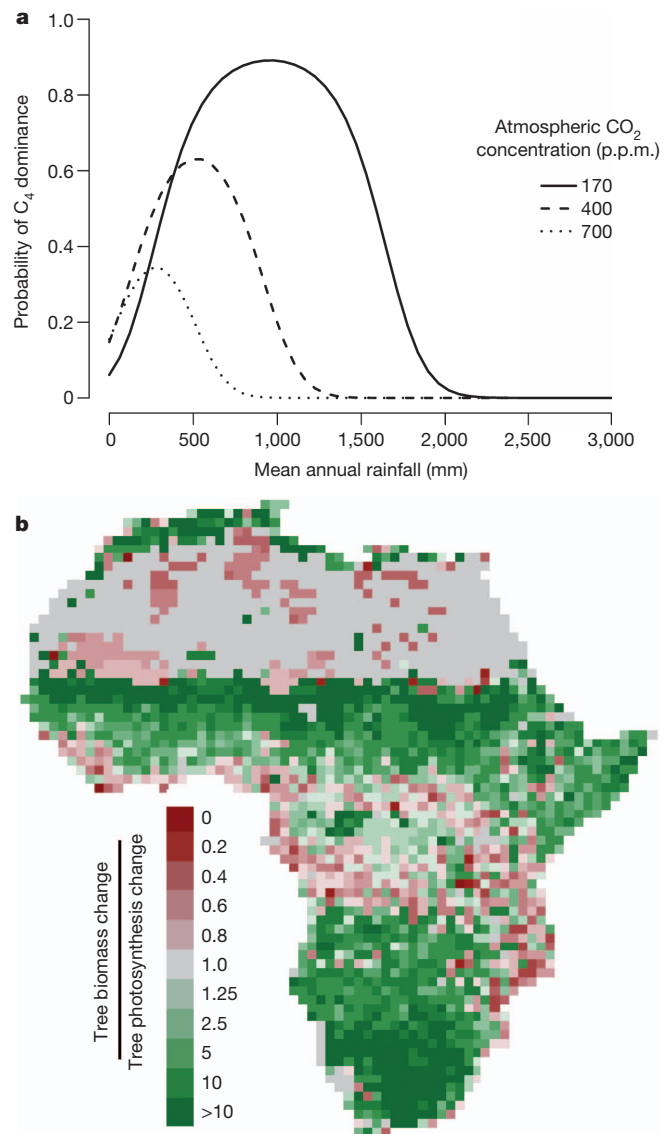


Figure 2 | Sensitivity to rainfall and atmospheric CO₂ concentration. **a**, The probability that a site will be C₄ dominated (C₄ grassland or savanna) as a function of rainfall for different atmospheric CO₂ concentrations. The data are from simulation runs for all sites in Africa. The low probability of C₄-dominated ecosystems at low rainfall is because deserts are more likely at low rainfall, whereas at high rainfall C₃-dominated ecosystems are more likely. **b**, The sensitivity of tree biomass change (the ratio of tree biomass in 2100 to that in 1850) to tree photosynthesis change (the ratio of tree photosynthesis in 2100 to that in 1850).

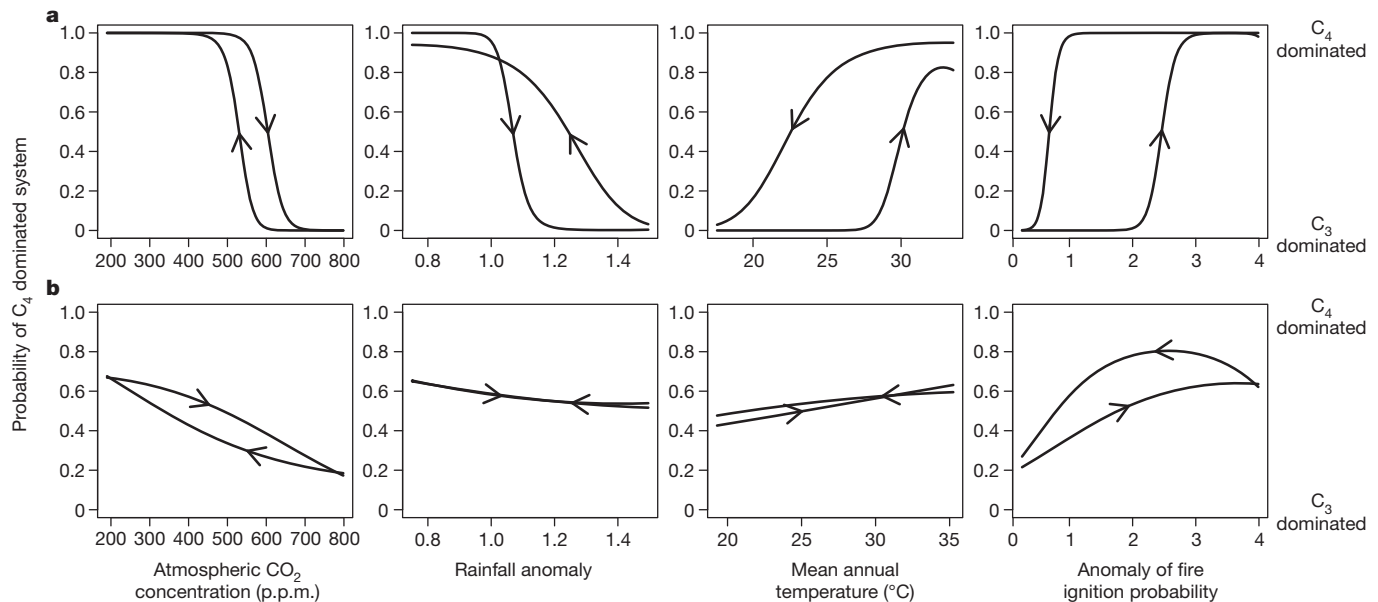


Figure 3 | Hysteresis in vegetation state. The probability that a site will be C_4 -dominated (C_4 grassland or savanna) as opposed to C_3 -dominated (evergreen forest, deciduous forest or woodland) when forced by atmospheric CO_2 , rainfall, mean annual temperature and fire ignitions. The arrows indicate whether the simulations were initiated with a low level of the forcing variable,

followed by an increase in the forcing variable or the converse (high initiation level, followed by a decrease). **a**, Local scale: responses estimated from 100 replicate simulations performed at a single site. **b**, Continental scale: estimates based on simulations replicated across Africa (excluding deserts and C_3 grasslands).

theory that the water use efficiency of vegetation increases with atmospheric CO_2 concentrations²². At the high end of the rainfall gradient (Fig. 2a), the shift to C_3 -dominated states is both forced and catalysed by atmospheric CO_2 . At pre-industrial CO_2 concentrations, C_4 dominance extends to regions with high rainfall (more than 1,500 mm), whereas at the CO_2 concentrations expected in 2100, C_4 -dominated states are restricted to regions with lower rainfall (less than 750 mm; Fig. 2a).

The trend towards increased woody biomass (Fig. 1a) is not a simple function of CO_2 concentration; it is a function of the conditions under which CO_2 fertilization shifts the competitive balance in favour of trees, setting up a positive feedback cycle involving light competition and fire suppression that further suppresses grasses and promotes trees^{8,9,17}. This is illustrated by the large between-site differences in the sensitivity of change in tree biomass to changes in tree photosynthetic rate (Fig. 2b). That is, although small changes in photosynthetic rates can translate into large changes in woody biomass, this sensitivity is highly dependent on context.

Hysteresis, the dependence of a system's response to changes in forcing factors on its history, is an indicator of systems characterized by catastrophic regime shifts⁶. There was clear evidence of a hysteresis effect in the transition between C_3 -dominated and C_4 -dominated vegetation states at local scales (Fig. 3a). That is, the critical level of the environmental factor (be it CO_2 concentration, rainfall, temperature or fire ignitions) required to force a transition to an alternative state was dependent on the system's historical states. When the results of the hysteresis simulations are averaged over the African continent, the hysteresis signal is removed or considerably weakened (Fig. 3b). This is simply because each site has its own critical threshold that is determined by site-specific environmental covariates; averaging over these covariates removes the hysteresis signal. The implication is that although abrupt regime shifts may be observed at local sites (Fig. 3a), these shifts will, when averaged over a continent, seem to occur smoothly (Fig. 3b). The consequence of this averaging effect is that any impulse to the atmosphere arising from the terrestrial land surface of the savanna complex may be smoothed.

Our findings that shifts from C_4 -dominated to C_3 -dominated ecosystems are to be expected in the savanna complex are consistent

with empirical observations^{23–25}, with chamber experiments on the response of savanna trees and C_4 grasses to atmospheric CO_2 concentration²⁶, and with leaf physiological models¹⁶. Although these projections of increasing tree dominance seem to contradict studies that project forest dieback in the Amazon¹², this may simply be because our analysis assumes that rainfall remains constant, whereas projections of Amazon forest dieback are based on climate simulations that project decreased rainfall¹⁴. The high uncertainty in precipitation change over Africa led us to assume that rainfall remained at ambient levels (see Supplementary Information), but future studies should consider changes in rainfall as well as feedbacks between vegetation and rainfall¹³.

Here we have shown that although abrupt shifts in vegetation states are possible at the local scale, the timing of these shifts will not be coincident at the continental scale. The implication is that although a considerable proportion of the African continent will—land use allowing—shift to more woody vegetation states, these continental changes will be spread over a longer period than local scale analyses suggest. This study, together with palaeoecological evidence²⁷, suggest that atmospheric CO_2 has been and will be a major factor shaping vegetation change.

METHODS SUMMARY

We used the aDGVM (adaptive Dynamic Global Vegetation Model), a dynamic global vegetation model²⁰ (see Supplementary Information). The aDGVM integrates plant physiological processes generally used in dynamic global vegetation models (DGVMs²⁸) with novel submodels that allow plants to dynamically adjust carbon allocation and leaf phenology to environmental conditions. The model considers four basic functional types: a savanna tree, a forest tree, a C_4 grass and a C_3 grass. We used a simplified dichotomous scheme to classify model output into seven vegetation types. We identified areas of vegetation change by running forward simulations for Africa at a 1° grid resolution for the period between 1850 and 2100 under Intergovernmental Panel on Climate Change (IPCC) Special Report on Emissions Scenarios (SRES) A1B projection provided by ECHAM5 (ref. 29). The large uncertainty in precipitation projections for Africa led us to assume that rainfall remained at ambient levels³⁰ (see Supplementary Information). The simulations focus on potential vegetation; in other words, land-use impacts are ignored. Testing whether the model system exhibited hysteresis effects was achieved by first increasing an environmental factor linearly from a low level to a high level and by decreasing the environmental factor linearly from a high level to a low level.

Full Methods and any associated references are available in the online version of the paper.

Received 5 January; accepted 17 May 2012.

Published online 27 June; corrected online 8 August 2012 (see full-text HTML version for details).

- Lenton, T. M. *et al.* Tipping elements in the Earth's climate system. *Proc. Natl Acad. Sci. USA* **105**, 1786–1793 (2008).
- Clark, J. S. *et al.* Ecological forecasts: an emerging imperative. *Science* **293**, 657–660 (2001).
- Lenton, T. M. Early warning of climate tipping points. *Nature Clim. Change* **1**, 201–209 (2011).
- Scheffer, M., Carpenter, S., Foley, J. A., Folke, C. & Walker, B. Catastrophic shifts in ecosystems. *Nature* **413**, 591–596 (2001).
- Carpenter, S. R. *et al.* Resilience and resistance of a lake phosphorus cycle before and after food web manipulation. *Am. Nat.* **140**, 781–798 (1992).
- Scheffer, M. & Carpenter, S. R. Catastrophic regime shifts in ecosystems: linking theory to observation. *Trends Ecol. Evol.* **18**, 648–656 (2003).
- Liu, Z., Wang, Y., Gallimore, R., Notaro, M. & Prentice, I. C. On the cause of abrupt vegetation collapse in north Africa during the Holocene: climate variability vs. vegetation feedback. *Geophys. Res. Lett.* **33**, L22709 (2006).
- Scheiter, S. & Higgins, S. I. Partitioning of root and shoot competition and the stability of savannas. *Am. Nat.* **170**, 587–601 (2007).
- Staver, A. C., Archibald, S. & Levin, S. A. The global extent and determinants of savanna and forest as alternative biome states. *Science* **334**, 230–232 (2011).
- Hirota, M., Holmgren, M., Van Nes, E. H. & Scheffer, M. Global resilience of tropical forest and savanna to critical transitions. *Science* **334**, 232–235 (2011).
- Walter, H. *Ecology of Tropical and Subtropical Vegetation* (Oliver & Boyd, 1971).
- Cox, P. M. *et al.* Amazonian forest dieback under climate-carbon cycle projections for the 21st century. *Theor. Appl. Climatol.* **78**, 137–156 (2004).
- Malhi, Y. *et al.* Exploring the likelihood and mechanism of a climate-change-induced dieback of the Amazon rainforest. *Proc. Natl Acad. Sci. USA* **106**, 20610–20615 (2009).
- Rammig, A. *et al.* Estimating the risk of Amazonian forest dieback. *New Phytol.* **187**, 694–706 (2010).
- Scholes, R. J. & Archer, S. R. Tree–grass interactions in savannas. *Annu. Rev. Ecol. Syst.* **28**, 517–544 (1997).
- Ehleringer, J. R., Cerling, T. E. & Helliker, B. R. C₄ photosynthesis, atmospheric CO₂ and climate. *Oecologia* **112**, 285–299 (1997).
- Bond, W. J. & Midgley, G. F. A proposed CO₂-controlled mechanism of woody plant invasion in grasslands and savannas. *Glob. Change Biol.* **6**, 865–869 (2000).
- Higgins, S. I., Bond, W. J. & Trollope, W. S. W. Fire, resprouting and variability: a recipe for grass–tree coexistence in savanna. *J. Ecol.* **88**, 213–229 (2000).
- Bond, W. J. What limits trees in C₄ grasslands and savannas? *Annu. Rev. Ecol. Evol. Syst.* **39**, 641–659 (2008).
- Scheiter, S. & Higgins, S. I. Impacts of climate change on the vegetation of Africa: an adaptive dynamic vegetation modelling approach. *Glob. Change Biol.* **15**, 2224–2246 (2009).
- Peel, M. C., Finlayson, B. L. & McMahon, T. A. Updated world map of the Köppen–Geiger climate classification. *Hydrol. Earth Syst. Sci.* **11**, 1633–1644 (2007).
- Drake, B. G., Gonzalez-Meler, M. A. & Long, S. P. More efficient plants: a consequence of rising atmospheric CO₂? *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48**, 609–639 (1997).
- Bowman, D. M. J. S., Murphy, B. P. & Banfai, D. S. Has global environmental change caused monsoon rainforests to expand in the Australian monsoon tropics? *Landscape Ecol.* **25**, 1247–1260 (2010).
- Wigley, B. J., Bond, W. J. & Hoffman, M. T. Thicket expansion in a South African savanna under divergent land use: local vs. global drivers? *Glob. Change Biol.* **16**, 964–976 (2010).
- Buitenwerf, R., Bond, W. J., Stevens, N. & Trollope, W. S. W. Increased tree densities in South African savannas: >50 years of data suggests CO₂ as a driver. *Glob. Change Biol.* **18**, 675–684 (2012).
- Kgope, B. S., Bond, W. J. & Midgley, G. F. Growth responses of African savanna trees implicate atmospheric CO₂ as a driver of past and current changes in savanna tree cover. *Austral Ecol.* **35**, 451–463 (2010).
- Prentice, A. I., Harrison, S. P. & Bartlein, P. J. Global vegetation and terrestrial carbon cycle changes after the last ice age. *New Phytol.* **189**, 988–998 (2011).
- Prentice, I. C. *et al.* in *Terrestrial Ecosystems in a Changing World* (eds Canadell, J. G., Pataki, D. & Pitelka, L. F.) 175–192 (Springer, 2007).
- Roeckner, E. IPCC DDC AR4 ECHAM5/MPI-OM SRESA1B run1. World Data Center for Climate, CERA-DB 'EH5 MPI OM SRESA1B 1' (http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=EH5_MPI_OM_SRESA1B_1) (2005).
- New, M., Lister, D., Hulme, M. & Makin, I. A high-resolution data set of surface climate over global land areas. *Clim. Res.* **21**, 1–25 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Benu and T. Hickler for helping us improve the manuscript. Financial support was provided by the Deutsche Forschungsgemeinschaft and Hesse's Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz (LOEWE) programme.

Author Contributions S.I.H. and S.S. jointly conceived the study. S.S. ran the simulations. S.I.H. led the writing.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.I.H. (higgins@em.uni-frankfurt.de).

METHODS

Basic model structure. We used the aDGVM (adaptive Dynamic Global Vegetation Model), a dynamic global vegetation model²⁰. A complete description is provided in ref. 20; we provide only a summary here. The aDGVM integrates plant physiological processes generally used in dynamic global vegetation models (DGVMs²⁸) with novel submodels that allow plants to dynamically adjust carbon allocation and leaf phenology to environmental conditions. The aDGVM is individual-based, which means that it keeps track of state variables such as above-ground and below-ground biomass, height, leaf area index and leaf phenological status for individual plants. The model considers four basic functional types: a savanna tree, a forest tree, a C_4 grass and a C_3 grass (see Supplementary Information). For this study we used a simplified dichotomous scheme to classify model output into seven vegetation types (Supplementary Fig. 1). **Simulation experiments.** We simulated the potential vegetation of Africa at a 1° grid resolution for the period between 1850 and 2100 under Intergovernmental Panel on Climate Change (IPCC) Special Report on Emissions Scenarios (SRES) A1B projection provided by ECHAM5 (ref. 29). The use of other scenarios does not qualitatively change our results. Because of the large uncertainty in the precipitation projections of the IPCC SRES simulations we assume that precipitation remains as defined by the Climate Research Unit's empirical climate data³⁰. For all simulations, the model was initialized with 100-year spin-up using the environmental conditions of 1850.

Forward simulations. We ran forward simulations from 1850 to 2100. Nine replicate simulations with and without fire were performed. Each simulated grid cell was classified into one of the seven vegetation types (Supplementary Fig. 1). There was high spatial agreement between the maps produced by replicate simulations ($\kappa > 0.75$). We calculated the transition rates between the different vegetation types observed between 1850 and 2100 (Fig. 1a) in the presence of fire. We plotted, for the most frequent transition categories, where these transitions occur in Africa (Fig. 1b). We plotted the distribution of bistable sites (C_4 grasslands and savannas that could be C_3 -dominated (woodland, deciduous forest or evergreen forest) if fire were suppressed) for 1850 and 2100 (Fig. 1c).

We ran 100 replicate simulations for each grid cell and estimated the year in which transitions from C_4 -dominated systems (C_4 grassland and savanna) to C_3 -dominated systems (woodland, deciduous forest and evergreen forest) occurred (Fig. 1d). Stable estimates of the mean time of transition for grid cells were obtained for sample sizes of 50 replicate simulations. The 95% confidence interval of the year in which a transition was observed in an individual grid cell spanned 48 years. Linear regression models were used to identify which factors influenced the mean time at which a transition occurred.

Sensitivity to CO_2 and rainfall. Simulations were run for constant CO_2 concentrations (170, 400 and 700 p.p.m.) and ambient climatic conditions for all Africa. These simulations were summarized by using a logistic regression to estimate how rainfall influenced the probability of vegetation being in a C_4 -dominated state (C_4 grassland and savanna; Fig. 2a).

The forward simulations were used to calculate the sensitivity of tree biomass change to changes in the photosynthetic rates of trees (Fig. 2b). This sensitivity was defined as (tree biomass at 2100/tree biomass at 1850)/(tree photosynthesis at 2100/tree photosynthesis at 1850); high sensitivity would imply that a small change in the photosynthetic rate of trees produces a large change in tree biomass. The photosynthetic rates are the potential photosynthetic rates (assuming saturating moisture supply, but variation in CO_2 , incoming photosynthetically active radiation and temperature with site)³⁰.

Hysteresis effects. Simulation experiments were conducted to explore whether the model showed hysteresis effects when vegetation underwent a transition between C_4 -dominated (C_4 grassland and savanna) and C_3 -dominated (woodland, deciduous forest and evergreen forest) states. We set all environmental conditions to ambient³⁰ conditions. For selected environmental factors (atmospheric CO_2 concentration, mean annual temperature anomaly, mean annual precipitation anomaly and fire ignition anomaly) we first increased the environmental factor linearly from a low level to a high level and then decreased the environmental factor linearly from a high level to a low level. The low and high levels for CO_2 were 190 p.p.m. and 800 p.p.m., respectively. The temperature anomaly was $-8^\circ C$ to $+8^\circ C$ relative to ambient. The mean annual precipitation anomaly was 0.75–1.5 of ambient. The fire ignition anomaly was 0–4 of the standard ignition probability. The period over which we increased and decreased the forcing variable was 350 years. Spin-up for these simulations involved running the model for 100 years at the high or low level of the forcing variable, with other parameters being set to ambient.

In a first hysteresis experiment we ran 100 replicate simulations for a single site ($25^\circ S$, $31^\circ 35' E$; 570 mm mean annual precipitation). For these simulations CO_2 was set to 550 p.p.m. because this was, for this site, the CO_2 concentration required for shifts from savanna to woodland and/or forest. At the continental scale, we simulated vegetation of Africa at a 2° grid resolution. In each simulation year we recorded whether each replicate simulation was in a C_4 -dominated or a C_3 -dominated state. For the local scale simulations these replicates were 100 repeat simulations at a single site. For the continental scale simulations the replicates were each pixel in the 2° grid of Africa. These data were then summarized using a logistic regression, with C_4 -dominated or C_3 -dominated as the binary response variable, and the forcing factor (atmospheric CO_2 concentration, mean annual temperature anomaly, mean annual precipitation anomaly or fire ignition anomaly) as the independent variable (Fig. 3).

The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants

Angélique D'Hont^{1*}, France Denoeud^{2,3,4*}, Jean-Marc Aury², Franc-Christophe Baurens¹, Françoise Carreel^{1,5}, Olivier Garsmeur¹, Benjamin Noel², Stéphanie Bocs¹, Gaëtan Droc¹, Mathieu Rouard⁶, Corinne Da Silva², Kamel Jabbari^{2,3,4}, Céline Cardil¹, Julie Poulain², Marlène Souquet¹, Karine Labadie², Cyril Jourda¹, Juliette Lengellé¹, Marguerite Rodier-Goud¹, Adriana Alberti², Maria Bernard², Margot Correa², Saravanaraj Ayyampalayam⁷, Michael R. Mckain⁷, Jim Leebens-Mack⁷, Diane Burgess⁸, Mike Freeling⁸, Didier Mbéguié-A-Mbéguié⁹, Matthieu Chabannes⁵, Thomas Wicker¹⁰, Olivier Panaud¹¹, Jose Barbosa¹¹, Eva Hribova¹², Pat Heslop-Harrison¹³, Rémy Habas⁵, Ronan Rivallan¹, Philippe Francois¹, Claire Poirion¹, Andrzej Kilian¹⁴, Dheema Burthia¹, Christophe Jenny¹, Frédéric Bakry¹, Spencer Brown¹⁵, Valentin Guignon^{1,6}, Gert Kema¹⁶, Miguel Dita¹⁹, Cees Waalwijk¹⁶, Steeve Joseph¹, Anne Dievart¹, Olivier Jaillon^{2,3,4}, Julie Leclercq¹, Xavier Argout¹, Eric Lyons¹⁷, Ana Almeida⁸, Mouna Jeridi¹, Jaroslav Dolezel¹², Nicolas Roux⁶, Ange-Marie Risterucci¹, Jean Weissenbach^{2,3,4}, Manuel Ruiz¹, Jean-Christophe Glaszmann¹, Francis Quétier¹⁸, Nabila Yahiaoui¹ & Patrick Wincker^{2,3,4}

Bananas (*Musa* spp.), including dessert and cooking types, are giant perennial monocotyledonous herbs of the order Zingiberales, a sister group to the well-studied Poales, which include cereals. Bananas are vital for food security in many tropical and subtropical countries and the most popular fruit in industrialized countries¹. The *Musa* domestication process started some 7,000 years ago in Southeast Asia. It involved hybridizations between diverse species and subspecies, fostered by human migrations², and selection of diploid and triploid seedless, parthenocarpic hybrids thereafter widely dispersed by vegetative propagation. Half of the current production relies on somaclones derived from a single triploid genotype (Cavendish)¹. Pests and diseases have gradually become adapted, representing an imminent danger for global banana production^{3,4}. Here we describe the draft sequence of the 523-megabase genome of a *Musa acuminata* doubled-haploid genotype, providing a crucial stepping-stone for genetic improvement of banana. We detected three rounds of whole-genome duplications in the *Musa* lineage, independently of those previously described in the Poales lineage and the one we detected in the Arecales lineage. This first monocotyledon high-continuity whole-genome sequence reported outside Poales represents an essential bridge for comparative genome analysis in plants. As such, it clarifies commelinid-monocotyledon phylogenetic relationships, reveals Poaceae-specific features and has led to the discovery of conserved non-coding sequences predating monocotyledon-eudicotyledon divergence.

Banana cultivars mainly involve *M. acuminata* (A genome) and *Musa balbisiana* (B genome) and are sometimes diploid but generally triploid^{5,6}. We sequenced the genome of DH-Pahang, a doubled-haploid *M. acuminata* genotype ($2n = 22$), of the subspecies *malaccensis* that contributed one of the three *acuminata* genomes of Cavendish⁷. A total of 27.5 million Roche/454 single reads and 2.1 million Sanger reads were produced, representing $20.5\times$ coverage of the 523-megabase (Mb) DH-Pahang genome size, as estimated by flow cytometry. In addition, $50\times$ of Illumina data were used to correct

sequence errors. The assembly consisted of 24,425 contigs and 7,513 scaffolds with a total length of 472.2 Mb, which represented 90% of the estimated DH-Pahang genome size. Ninety per cent of the assembly was in 647 scaffolds, and the N50 (the scaffold size above which 50% of the total length of the sequence assembly can be found) was 1.3 Mb (Supplementary Text and Supplementary Tables 1–3). We anchored 70% of the assembly (332 Mb) along the 11 *Musa* linkage groups of the Pahang genetic map. This corresponded to 258 scaffolds and included 98.0% of the scaffolds larger than 1 Mb and 92% of the annotated genes (Supplementary Text, Supplementary Table 4 and Supplementary Fig. 1).

We identified 36,542 protein-coding gene models in the *Musa* genome (Supplementary Tables 1 and 5). A total of 235 microRNAs from 37 families were identified, including only one of the eight microRNA gene (*MIR*) families found so far solely in Poaceae⁸ (Supplementary Tables 6 and 7).

Viral sequences related to the banana streak virus (BSV) dsDNA plant pararetrovirus were found to be integrated in the Pahang genome, with 24 loci spanning 10 chromosomes (Supplementary Text and Supplementary Fig. 2). They belonged to a badnavirus phylogenetic group that differed from the endogenous BSV species (eBSV) found in *M. balbisiana*⁹ and most of them formed a new subgroup (Supplementary Fig. 3). Importantly, all of the integrations were highly reorganized and fragmented and thus did not seem to be capable of forming free infectious viral particles, contrary to the eBSV described in *M. balbisiana*¹⁰.

Transposable elements account for almost half of the *Musa* sequence (Supplementary Text and Supplementary Tables 1 and 8–10). Long terminal repeat retrotransposons represent the largest part, with *Copia* elements being much more abundant than *Gypsy* elements (25.7–11.6%) (Supplementary Fig. 4). No major recent wave of long terminal repeat retrotransposon insertions appears to have occurred in the *Musa* lineage. Fewer than 1% of the long terminal repeat retrotransposons are complete and their median date of insertion is around 4 Myr ago, corresponding to the half-life of this type of

¹Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), UMR AGAP, F-34398 Montpellier, France. ²Commissariat à l'Energie Atomique (CEA), Institut de Génétique (IG), Genoscope, 2 rue Gaston Crémieux, BP5706, 91057 Evry, France. ³Centre National de Recherche Scientifique (CNRS), UMR 8030, CP5706, Evry, France. ⁴Université d'Evry, UMR 8030, CP5706, Evry, France. ⁵CIRAD, UMR BGPI, Campus international de Baillarguet, F-34398 Montpellier, France. ⁶Bioversity International, Parc Scientifique Agropolis II, 34397 Montpellier Cedex 5, France. ⁷Department of Plant Biology, University of Georgia, Athens, Georgia 30602, USA. ⁸Department of Plant and Microbial Biology, University of California, Berkeley, California 94720, USA. ⁹CIRAD, UMR QUALISUD Station de Neufchâteau, Sainte-Marie, 97130 Capesterre-Belle-Eau, France. ¹⁰Institute of Plant Biology, University of Zurich, CH-8008 Zurich, Switzerland. ¹¹Laboratoire Génome et Développement des Plantes, UMR 5096 CNRS-UPVD, 66000 Perpignan, France. ¹²Centre of the Region Hana for Biotechnological and Agricultural Research, Institute of Experimental Botany, Sokolovska 6, CZ-77200 Olomouc, Czech Republic. ¹³Department of Biology, University of Leicester, Leicester LE1 7RH, UK. ¹⁴Diversity Arrays Technology, Yarralumla, Australian Capital Territory 2600, Australia. ¹⁵Institut des Sciences du Végétal, CNRS UPR 2355 et FRC3115, 91198 Gif-sur-Yvette, France. ¹⁶University of Wageningen, Plant Research International, 6700 AA Wageningen, Netherlands. ¹⁷Department of Plant Sciences, University of Arizona, Tucson, Arizona, USA. ¹⁸Département de Biologie, Université d'Evry Val d'Essonne, Evry, France. ¹⁹Brazilian Agricultural Research Corporation (EMBRAPA), Embrapa Cassava & Fruits, Cruz das Almas, 44380-000, Salvador, Bahia, Brazil.

*These authors contributed equally to this work.

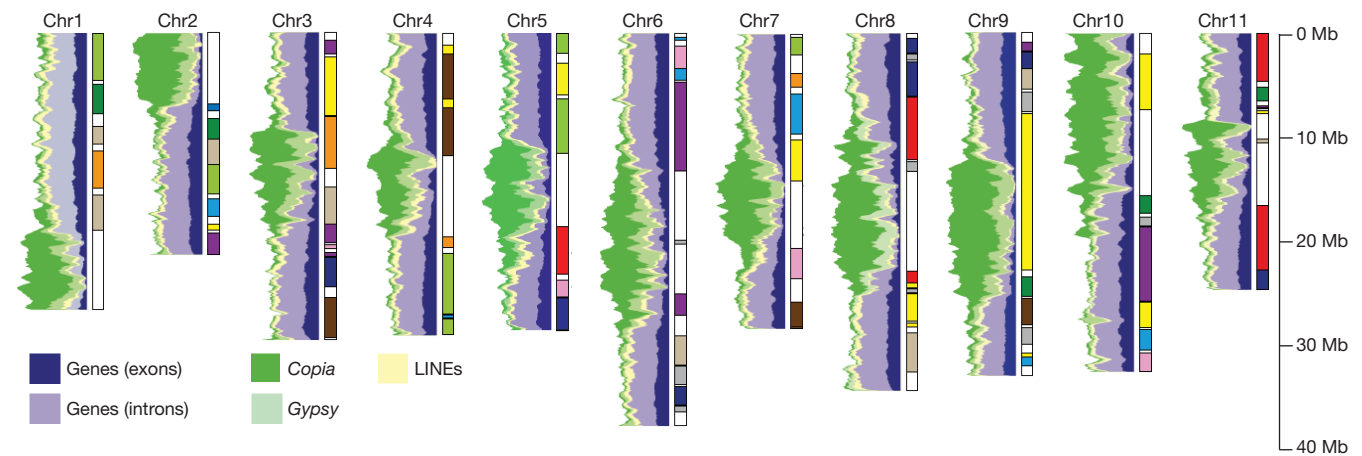


Figure 1 | Chromosomal distribution of the main *M. acuminata* genome features. Distribution of genes and transposable elements (left) and paralogous relationships between the 11 chromosomes indicated with 12 distinct colours

transposable element¹¹ (Supplementary Fig. 5). Long interspersed elements (LINEs) represent 5.5% of the genome. The banana genome is exceptional in the composition of its class 2 element population, which represents only about 1.3% of the genome. The only superfamilies identified were *hAT*, followed by *Harbinger* and *Mutator*. Only the first family was significantly represented and had non-autonomous deletion derivatives. The superfamilies *CACTA* and *Mariner*, which have been found in high copy numbers in all angiosperm genomes studied so far, are absent from the banana genome. Gene-rich regions are mostly located on distal parts of chromosomes, as observed in other plant genomes (Fig. 1 and Supplementary Fig. 1). There is, however, a particularly sharp transition between gene-rich and transposable-element-rich regions. This observation is confirmed by the pattern observed after genomic *in situ* hybridization, which shows that transposable elements are typically concentrated around centromeres in *Musa*¹² (Supplementary Fig. 6). The asymmetric transposable element distributions along the chromosomes indicated that chromosomes 1 and 2 are acrocentric in DH-Pahang (Fig. 1). Long terminal repeat retrotransposons are particularly abundant in centromeric and pericentromeric chromosome regions. Their accumulation in these regions, particularly for the oldest ones, suggests that they are preferentially eliminated from gene-rich regions¹³ (Supplementary Fig. 5). Remarkably, typical short tandem centromeric repeats were not found in *Musa*. However, one long interspersed element (named *Nanica*) identified in the unassembled reads was localized by fluorescence *in situ* hybridization in the centromeric region of all *Musa* chromosomes (Supplementary Fig. 7 and Supplementary Table 10).

Whole-genome duplications (WGDs) have played a major role in angiosperm genome evolution¹⁴; the first evidence of a WGD event in the *Musa* lineage was reported by Lescot *et al.*¹⁵. We uncovered a complex pattern of paralogous relationships between the 11 *Musa* chromosomes (Supplementary Text and Supplementary Fig. 8). Most paralogous gene clusters shared relationships with three other clusters, suggesting that two WGDs (denoted as α and β) occurred (Supplementary Fig. 9). Based on *Ks* and synteny relationships, duplicated gene clusters were tentatively assembled into 12 *Musa* ancestral blocks representing the ancestral genome before the α/β duplications (Figs 1 and 2 and Supplementary Figs 10–12). The duplicated segments included in the *Musa* ancestral blocks cover 222 Mb (67% of the anchored assembly) and contain 26,829 genes (80% of the anchored genes) (Supplementary Table 11). The *Ks* distribution among pairs of paralogous gene clusters dated the two WGDs at a similar period around 65 Myr ago (Supplementary Fig. 13), consistent with the WGDs that occurred in many different plant

corresponding to the 12 *Musa* α/β ancestral blocks (right). LINEs, long interspersed elements.

lineages near the Cretaceous–Tertiary boundary¹⁴ (Fig. 3). Additional paralogous relationships between the 12 *Musa* ancestral blocks displaying higher *Ks* values suggested that an additional, more ancient duplication event (denoted as γ) occurred around 100 Myr ago (Fig. 3 and Supplementary Figs 10, 11, 13 and 14).

In the grass lineage, it is well established that one WGD (denoted as ρ) occurred around 50–70 Myr ago, after Poales separated from other monocotyledon orders^{16,17}. Evidence was reported on an additional WGD (denoted as σ) earlier in the monocotyledon lineage, but after its divergence from the eudicotyledons¹⁸. Our comparison of the *Musa* ancestral blocks with the Poaceae ρ and σ ancestral blocks as defined by Tang *et al.*¹⁸ revealed that genes from segments of different ρ blocks (corresponding to one σ block) have orthologous relationships with the same *Musa* regions, showing that the σ Poaceae event is not shared with *Musa*. Reciprocally, genes from *Musa* α/β paralogous segments

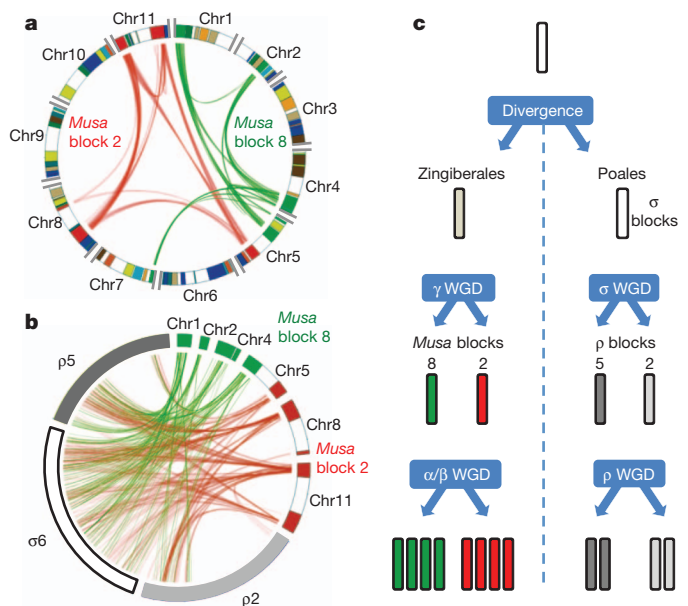


Figure 2 | Whole-genome duplication events. **a**, Paralogous relationships between chromosome segments from *Musa* α/β ancestral blocks 2 (red) and 8 (green). The 12 *Musa* α/β ancestral blocks are shown in different colours on the circle. **b**, Orthologous relationships of *Musa* ancestral blocks 2 and 8 with rice ancestral blocks ρ_2 , ρ_5 and ρ_6 . We did not observe a one-to-one relationship between, for instance, *Musa* α/β ancestral block 2 and one ρ ancestral block, which suggests that the γ and σ duplications are two separate events. **c**, Representation of the deduced WGD event.

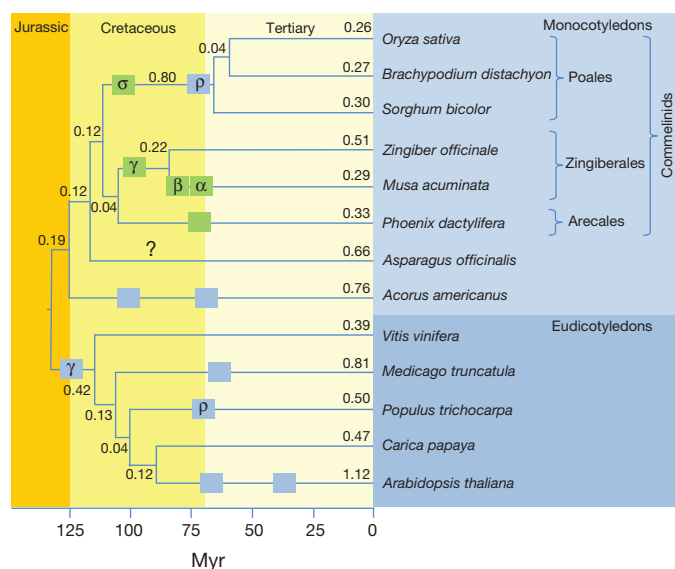


Figure 3 | Timing of whole-genome duplications relative to speciation events within representative monocotyledons and eudicotyledons. Boxes indicate WGD events. Green boxes indicate WGD events analysed in this paper. All nodes have 100% bootstrap support in a maximum likelihood analysis. Branch lengths (synonymous substitution rate) are indicated. The timing of the β WGD event relative to the Musaceae/Zingiberaceae split remains to be clarified.

have orthologous relationships with the same ρ and σ regions, showing that the earliest duplication (γ) we identified in the *Musa* lineage is not shared with Poaceae (Fig. 2 and Supplementary Fig. 15).

Independent phylogenomic analyses performed on 3,553 gene families, including genes mapped to syntenic ancestral blocks, generated further evidence (98.7–77.6% of the gene trees, Supplementary Text) that the three rounds of palaeopolyploidization identified in the *Musa* genome and the two previously reported in the Poaceae lineage occurred independently after the Poales and Zingiberales divergence estimated at 109–123 Myr ago¹⁹ (Fig. 3 and Supplementary Fig. 16).

Resolution of the Zingiberales relationship relative to Poales and Arecales (palms) has been problematic (see, for example, Givnish *et al.*²⁰), but our analysis of 93 single-copy nuclear genes suggested that the palms are more closely related to Zingiberales (including *Musa*) than to Poales (Fig. 3, Supplementary Text and Supplementary Fig. 17). Phylogenomic and synteny analyses indicated that the palms do not, however, share any of the Poales or Zingiberales WGDs discussed here (Supplementary Figs 17 and 18). Moreover, our *Ks* analyses of date-palm gene models²¹ indicated that the palm genome had its own WGD event (Supplementary Fig. 19).

Most (65.4%) of the genes included in the *Musa* α/β ancestral blocks are singletons and only 10% are retained in four copies, in agreement with the loss of most gene-duplicated copies after WGD²². The most retained gene ontology categories corresponded to genes involved in transcription regulation (transcription factor activity), signal transduction including small GTPase-mediated signal transduction and protein kinases, and translational elongation (Supplementary Text and Supplementary Tables 12–14). This might be explained by the gene balance hypothesis²³, which suggests that genes involved in multi-protein complexes or regulatory genes are dosage sensitive and thus are more prone to be co-retained or co-lost after WGD²⁴. With 3,155 genes, the number of *Musa* transcription factors identified is among the highest of all sequenced plant genomes (Supplementary Table 15 and 16).

Comparison of *Musa*, rice, sorghum, *Brachypodium*, date palm (*Phoenix dactylifera*) and *Arabidopsis* proteomes revealed 7,674 gene clusters in common to all six species, thus representing ancestral gene

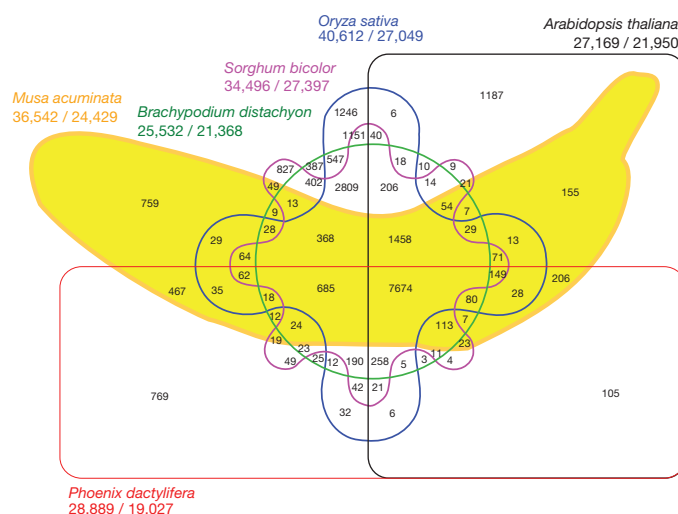


Figure 4 | Six-way Venn diagram showing the distribution of shared gene families (sequence clusters) among *M. acuminata*, *P. dactylifera*, *Arabidopsis thaliana*, *Oryza sativa*, *Sorghum bicolor* and *Brachypodium distachyon* genomes. Numbers of clusters are provided in the intersections. The total number of sequences for each species is provided under the species name (total number of sequences/total number of clustered sequences).

families (Fig. 4). Interestingly, many specific clusters (2,809 in our setting) proved specific to Poaceae, suggesting a high level of gene divergence and diversification within the grass lineage. Specific *Musa* clusters (759) were enriched in genes encoding transcription factors (for example, *Myb* and *AP2/ERF* families), defence-related proteins, enzymes of cell-wall biosynthesis and enzymes of secondary metabolism (Supplementary Table 17).

We compared the distribution of GC3 content (G or C in the third codon position) in *Musa* coding sequences with those of rice, ginger (*Zingiber officinale*) and date palm because this distribution was shown to be bimodal in Poaceae and unimodal in all analysed eudicotyledons²⁵. In *Musa*, a GC-rich peak was apparent but less distinct from the GC-poor one (Supplementary Text, Supplementary Figs 20–23 and Supplementary Table 18), which confirms preliminary evidence that placed *Musa* in an intermediate position¹⁵. This feature was shared with ginger (Zingiberales) and contrasts with the unimodal GC distribution of date-palm coding sequences (Supplementary Fig. 21).

Plant conserved non-coding sequences (CNSs)—a type of phylogenetic footprint—are enriched in known transcription factors or other *cis*-acting binding sites, and are usually clustered around regulatory genes, supporting their functionality²⁶. Starting with a collection of 16,978 CNSs conserved in Poaceae, we used the *Musa* genome to identify the 116 most deeply conserved regulatory binding sequences in the commelinid monocotyledon lineage (Supplementary Text, Supplementary Tables 19 and 20, and Supplementary Fig. 24). Deeply conserved CNSs in commelinids were frequently found located 5' to genes encoding transcription factors, and were significantly enriched in WRKY motifs (Supplementary Table 21). After WGD, genes associated with deeply conserved CNSs were found to be retained as duplicates more often than genes with less deeply conserved CNSs (Supplementary Table 22). The banana genome also served as a stepping-stone to finding CNSs conserved beyond monocotyledons, including 18 CNSs that were found in this study to be conserved in the expected syntenic position in eudicotyledons as well (Supplementary Table 23). This evolutionary distance is not unusual for vertebrate CNSs (detectable after more than 400 million years of divergence)²⁷ but it surpasses the findings of previous plant whole-genome surveys²⁶. Plant deeply conserved CNSs are therefore rare but do exist, and are short compared with those of animals²⁷, and must be at least as old as

monocotyledon–eudicotyledon divergence (more than 130 million years of divergence).

The reference *Musa* genome sequence represents a major advance in the quest to unravel the complex genetics of this vital crop, whose breeding is particularly challenging. Having access to the entire *Musa* gene repertoire is a key to identifying genes responsible for important agronomic characters, such as fruit quality and pest resistance. Bananas are exported green and then ripened by application of ethylene. RNA-Seq analysis indicated strong transcriptional reprogramming in mature green banana fruits after ethylenic treatment (Supplementary Text, Supplementary Tables 24–26 and Supplementary Fig. 25). Transcription factors were particularly involved with 597 differentially regulated genes. Various modifications confirmed the biochemistry of the banana ripening process²⁸, such as highly upregulated genes encoding cell-wall modifying enzymes, three downregulated starch synthase genes and one upregulated β -amylase gene. Two WGD-derived paralogous vacuolar invertase genes involved in sucrose conversion displayed opposite expression profiles, suggesting subfunctionalization and possible contribution to the soluble sugar balance in ripening bananas (Supplementary Fig. 26). The race against pathogen evolution is particularly critical in clonally propagated crops such as banana. Up to 50 pesticide treatments a year are required in large plantations against black leaf streak disease, a recent pandemic caused by *Mycosphaerella fijiensis*³. Moreover, outbreaks of a new race of the devastating Panama disease fungus (*Fusarium oxysporum* f. sp. *cubense*) are spreading in Asia⁴. Among defence-related genes, those encoding nucleotide-binding site leucine-rich repeat proteins were found to be little represented in the *Musa* sequence (89 genes) (Supplementary Table 27). RNA-Seq analysis showed that receptor-like kinase genes were upregulated in a partially resistant interaction with *M. fijiensis* (Supplementary Text, Supplementary Table 28 and Supplementary Fig. 27). Interestingly, direct links between basal plant immunity triggered by receptor-like kinase proteins and quantitative trait loci for partial resistance have been recently established in several plant species (see, for example, Poland *et al.*²⁹). In addition, we showed that DH-Pahang is highly resistant to the new broad-range *Fusarium oxysporum* race 4 (Supplementary Text and Supplementary Fig. 28), thus conferring additional specific value to the DH-Pahang sequence.

The *Musa* genome sequence reported here bridges a large gap in genome evolution studies. As such, it sheds new light on the monocotyledon lineage. Several Poaceae-specific characteristics could be highlighted, boosting prospects for analysing the emergence of this very successful family. The *Musa* genome also enabled identification of deeply conserved CNS within commelinid monocotyledons and between monocotyledons and eudicotyledons, representing an invaluable resource for detecting novel motifs with a gene regulation function. We detected three rounds of polyploidization in the *Musa* lineage, which were followed by gene loss and chromosome rearrangements, resulting in little synteny conservation between lineages (Supplementary Figs 29 and 30) and over-retention of some gene classes, thus providing ample opportunities for independent diversification. In particular, transcription factor families are strikingly expanded in *Musa* compared with other plant genomes and probably contribute to specific aspects of banana development.

The *Musa* genome sequence is therefore an important advance towards securing food supplies from new generations of *Musa* crops, and provides an invaluable stepping-stone for plant gene and genome evolution studies.

METHODS SUMMARY

Sanger (ABI 3730xl sequencers) and Roche/454 (GSFLX pyrosequencing platform) reads were assembled with Newbler. Scaffolds were anchored to Pahang linkage groups using 652 markers (SSR and DArT). Protein-coding gene model prediction on the repeat-masked sequence was done with the GAZE³⁰ computational framework by combining *ab initio* gene predictions, protein similarity, existing banana and monocotyledon transcript information and banana RNA-Seq data. A reference library of *Musa* transposable elements was built based

on sequence similarity at the protein and nucleic acid levels and on searches for transposable-element structural signatures. The library was used with the REPET package (<http://urgi.versailles.inra.fr/Tools/REPET>) to screen the *Musa* assembly and quantify repeats.

RNA-Seq differential gene expression analysis was performed using Illumina GAIIX 76 bases reads that were mapped to the DH-Pahang sequence using SOAP2 (<http://soap.genomics.org.cn/>).

Full Methods and any associated references are available in the online version of the paper.

Received 10 February; accepted 18 May 2012.

Published online 11 July; corrected online 8 August 2012 (see full-text HTML for details).

- Lescot, T. The genetic diversity of banana in figures. *Fruit Trop* **189**, 58–62 (2011).
- Perrier, X. *et al.* Multidisciplinary perspectives on banana (*Musa* spp.) domestication. *Proc. Natl Acad. Sci. USA* **108**, 11311–11318 (2011).
- De Lapeyre de Bellaire, L., Fouré, E., Abadie, C. & Carlier, J. Black leaf streak disease is challenging the banana industry. *Fruits* **65**, 327–342 (2010).
- Dita, M. A., Waalwijk, C., Buddenhagen, I. W., Souza, M. T. & Kema, G. H. J. A molecular diagnostic for tropical race 4 of the banana fusarium wilt pathogen. *Plant Pathol.* **59**, 348–357 (2010).
- Simmonds, N. W. *The Evolution of the Bananas* 101–131 (Longman, 1962).
- D'Hont, A., Paget-Goy, A., Escoute, J. & Carreel, F. The interspecific genome structure of cultivated banana, *Musa* spp. revealed by genomic DNA *in situ* hybridization. *Theor. Appl. Genet.* **100**, 177–183 (2000).
- Raboin, L. M. *et al.* Diploid ancestors of triploid export banana cultivars: molecular identification of 2n restitution gamete donors and n gamete donors. *Mol. Breed.* **16**, 333–341 (2005).
- Cuperus, J. T., Fahlgren, N. & Carrington, J. C. Evolution and functional diversification of MIRNA genes. *Plant Cell* **23**, 431–442 (2011).
- Gayral, P. & Iskra-Caruana, M. L. Phylogeny of banana streak virus reveals recent and repetitive endogenization in the genome of its banana host (*Musa* spp.). *J. Mol. Evol.* **69**, 65–80 (2009).
- Iskra-Caruana, M. L., Baurens, F. C., Gayral, P. & Chabannes, M. A four-partner plant-virus interaction: enemies can also come from within. *Mol. Plant Microbe Interact.* **23**, 1394–1402 (2010).
- Ma, J., Devos, K. M. & Bennetzen, J. L. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res.* **14**, 860–869 (2004).
- Jeridi, M. *et al.* Homoeologous chromosome pairing between the A and B genomes of *Musa* spp. revealed by genomic *in situ* hybridization. *Ann. Bot. (Lond.)* **108**, 975–981 (2011).
- Paterson, A. H. *et al.* The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457**, 551–556 (2009).
- Van de Peer, Y., Fawcett, J. A., Proost, S., Sterck, L. & Vandepoele, K. The flowering world: a tale of duplications. *Trends Plant Sci.* **14**, 680–688 (2009).
- Lescot, M. *et al.* Insights into the *Musa* genome: syntenic relationships to rice and between *Musa* species. *BMC Genomics* **9**, 58 (2008).
- Paterson, A. H. *et al.* Comparative genome analysis of monocots and dicots, toward characterization of angiosperm diversity. *Curr. Opin. Biotechnol.* **15**, 120–125 (2004).
- Salse, J. *et al.* Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. *Plant Cell* **20**, 11–24 (2008).
- Tang, H., Bowers, J. E., Wang, X. & Paterson, A. H. Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. *Proc. Natl Acad. Sci. USA* **107**, 472–477 (2010).
- Magallon, S. & Castillo, A. Angiosperm diversification through time. *Am. J. Bot.* **96**, 349–365 (2009).
- Givnish, T. J. *et al.* Assembling the tree of the monocotyledons: plastome sequence phylogeny and evolution of Poales. *Ann. Mo. Bot. Gard.* **97**, 584–616 (2010).
- Al-Dous, E. K. *et al.* De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nature Biotechnol.* **29**, 521–527 (2011).
- Schnable, J. C., Springer, N. M. & Freeling, M. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl Acad. Sci. USA* **108**, 4069–4074 (2011).
- Birchler, J. A., Riddle, N. C., Auger, D. L. & Veitia, R. A. Dosage balance in gene regulation: biological implications. *Trends Genet.* **21**, 219–226 (2005).
- Veitia, R. A., Bottani, S. & Birchler, J. A. Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends Genet.* **24**, 390–397 (2008).
- Carels, N. & Bernardi, G. Two classes of genes in plants. *Genetics* **154**, 1819–1825 (2000).
- Freeling, M. & Subramaniam, S. Conserved noncoding sequences (CNSs) in higher plants. *Curr. Opin. Plant Biol.* **12**, 126–132 (2009).
- Woolfe, A. *et al.* Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* **3**, e7 (2005).
- Fils-Lycaon, B. *et al.* Acid invertase as a serious candidate to control the balance sucrose versus (glucose plus fructose) of banana fruit during ripening. *Sci. Hortic. (Amsterdam)* **129**, 197–206 (2011).
- Poland, J. A., Bradbury, P. J., Buckler, E. S. & Nelson, R. J. Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. *Proc. Natl Acad. Sci. USA* **108**, 6893–6898 (2011).

30. Howe, K. L., Chothia, T. & Durbin, R. GAZE: a generic framework for the integration of gene-prediction data by dynamic programming. *Genome Res.* **12**, 1418–1427 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was mainly supported by French National Research Agency, Commissariat à l'Energie Atomique and Centre de coopération Internationale en Recherche Agronomique pour le Développement. The Generation Challenge program supported DArT genotyping, and Stichting Het Groene Woudt part of BAC-end sequencing. We thank M. Teixeira Souza for authorizing early access to BAC-End sequences, L. Baudouin and T. Hardcastle for their help with the Bayseq analysis, O. Inizan, T. Flutre and F. Choulet for their help in transposable-element mapping. We thank the SouthGreen Bioinformatics Platform (<http://southgreen.cirad.fr>) for providing us with computational resources. We thank D. Manley for his help with the English in this paper.

Author Contributions J.M.A., F.C.B., F.C., O.G., B.N. and S.B. contributed equally to this work. J.P., K.L., J.M.A. and M.B. performed sequencing and assembly. S.Br. performed genome size evaluation. F.C.B., N.R. and G.K. built or gave access to BAC libraries or BAC-end sequences. B.N., S.Bo., F.D., M.Co., J.Len., C.D.S., G.D., M.Ro., N.Y., F.C.B. and V.G. performed protein coding gene annotation. F.C., F.C.B., C.C., M.S., G.D., R.H., R.R., P.F., A.K., C.Je., F.B., S.J., M.R.G. and A.M.R. performed plant material development, ploidy analysis, DNA extraction, markers development, genotyping, genetic

mapping, anchoring. K.J. performed gene GC content analyses. S.Bo., O.G., T.W., E.H., P.H.H., J.B., M.R.G., D.Burt., A.D.H., M.J., C.P., J.D., O.P., J.Len., G.D. and N.Y. performed transposable-element analysis. O.G., F.D., A.D.H., J.M.A., G.D., F.C.B., E.L., S.Bo. and O.J. performed WGD analyses based on synteny conservation. J.L.M., S.A., M.R.M., A.D.H., O.G. performed phylogenomic analyses of WGD. N.Y., M.Ro., J.Len., S.Bo., C.Jo., A.D., F.D., M.Ru. and A.Alm. performed gene family analyses. J.Lec., X.A., G.D. and S.Bo. performed transfer RNA and microRNA analyses. F.C.B. and M.Ch. performed endogenous virus analyses. D.Burg. and M.F. performed CNS analyses. N.Y., C.Jo., C.D.S., A.Alb., F.C., D.M.M., M.D., C.W., G.K., M.S., performed RNA extraction, phenotyping and/or transcriptomic analyses. A.D.H., N.Y., O.G., F.C., F.C.B., F.D., J.M.A., J.C.G., P.W., S.Bo., F.Q. and J.W. wrote or revised the paper. A.D.H., N.Y. and P.W. conceived and coordinated the whole project.

Author Information The final assembly and annotation are deposited in DDBJ/EMBL/GenBank under accession numbers CAIC01000001–CAIC01024424 (contigs), HE806462–HE813974 (scaffolds) and HE813975–HE813985 (chromosomes). Genome sequence and annotation can be obtained and viewed at <http://banana-genome.cirad.fr>. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.D.H. (angelique.d'hont@cirad.fr) or P.W. (pwinker@genoscope.cns.fr).

METHODS

Plant material and DNA preparation. Doubled-haploid Pahang (DH-Pahang, ITC1511) was obtained from wild *M. acuminata* subspecies *malaccensis* accession 'Pahang' through anther culture and spontaneous chromosome doubling³¹. Genome sizes were estimated by flow cytometry according to Marie and Brown³². High molecular weight DNA was prepared from the youngest fully expanded leaf of DH-Pahang as described in Piffanelli *et al.*³³ with minor modifications (Supplementary Text).

Genome sequencing. The genome was sequenced using a Whole Genome Shotgun strategy combining Sanger, Roche/454 GSFLX and Illumina GAIIX technologies. Sanger sequencing was performed with the ABI 3730xl on 10-kilobase (kb) inserts and on two BAC libraries generated with the HindIII and BamHI restriction enzymes resulting in 2.0 million 10-kb fragment-ends and about 90,500 BAC-ends. A total of 27.5 million reads were obtained using Roche/454 GSFLX.

Genome assembly and automatic error corrections with Solexa/Illumina reads. All reads were assembled with Newbler version MapAsmResearch-03/15/2010. From the initial 29,620,875 reads, 87.8% were assembled. We obtained 24,425 contigs that were linked into 7,513 scaffolds. The contig N50 (the contig size above which 50% of the total length of the sequence assembly is included) was 43.1 kb, and the scaffold N50 was 1.3 Mb. The cumulative scaffold size was 472.2 Mb, about 10% smaller than the estimated genome size of 523 Mb. Sequence quality of scaffolds from the Newbler assembly was improved as described previously³⁴, by automatic error corrections with Solexa/Illumina reads (50-fold genome coverage), which have a different bias in error type compared with 454 reads. To validate the assembly, we built a unigene set corresponding to 15,017 isotigs that were obtained from the assembly with Newbler (version MapAsmResearch-03/15/2010) of Roche/454 GSFLX reads from six different complementary DNA (cDNA) libraries (829,587 reads, Supplementary Text). The unigenes were aligned with the assembly using the BLAT algorithm³⁵ with default parameters, and the best match was kept for each unigene. The assembly covers a very large proportion of the euchromatin of the *M. acuminata* genome, as 99% of the set of 15,017 unigenes was recovered in the DH-Pahang genome assembly.

Construction of the Pahang genetic map and sequence anchoring. A genetic map was specifically developed for scaffold anchoring and orientation. A total of 2,454 single sequence repeats (SSR) markers and 1,008 polymorphic diversity array technology (DArT) markers were analysed including 1,411 new SSRs defined on sequence contigs and scaffolds. The map used for anchoring was built with 589 SSR and 63 DArT markers that were genotyped on 180 individuals of the Pahang self progeny. Data were analysed using JoinMap 4 (Plant Research International). The 652 markers anchored 258 scaffolds along the 11 linkage groups of the genetic map. Orientation of scaffolds was possible when two or more separated genetic markers were present on the same scaffold. All these data were used to generate 11 banana pseudo-chromosomes with 100Ns inserted between neighbouring scaffolds (Supplementary Fig. 1 and Supplementary Table 4).

Gene prediction. The following resources were integrated to automatically build *Musa acuminata* gene models using GAZE³⁰: *ab initio* gene predictions from Geneid³⁶, SNAP³⁷ and FGENESH³⁸; Genewise³⁹ alignments of the UniProt⁴⁰ database; Est2genome⁴¹ alignments of full-length cDNAs from six tissue samples of DH-Pahang and a collection of 6,888,879 monocotyledon messenger RNAs from the EMBL database and finally Gmorse models⁴² derived from RNA-Seq reads (Supplementary Text). MicroRNAs were predicted based on comparison using the Plant MicroRNA Database (<http://bioinformatics.cau.edu.cn/PMRD/>).

Identification of integrated pararetrovirus sequences. Viral integrants in the DH-Pahang genome were detected with a BLASTN analysis using either full-length BSV sequences or a 540-base-pair fragment of the RT/RNase H region of the badnaviruses genome (Supplementary Text).

Identification, classification and distribution of *Musa* transposable elements. *Musa* transposable elements were identified based on sequence similarity at the protein and nucleic-acid levels using BLASTP and TBLASTN⁴³ and by *de novo* identification based on transposable-element structural signatures. Repeats from 1,832,094 remaining unassembled reads were characterized with a BLASTN 'walking' approach⁴⁴. The obtained reference *Musa* transposable-element library was used with REPET⁴⁵ to screen the assembly and quantify repeats (Supplementary Text). Insertion dates of full-length long terminal repeat retrotransposons were determined as described in Ma *et al.*⁴⁶ with a substitution rate of 9×10^{-9} per site per year, which is twofold higher than that determined for *Musa* genes by Lescot *et al.*¹⁵.

Identification of *Musa* WGDs and comparative genome analyses. For the identification of *Musa* WGD, an all-against-all comparison of *Musa* proteins was done using the GenomeQuest BLAST package (LASSAP⁴⁷) and retaining ten best hits for each gene. Clusters of paralogues composed of at least 20 genes with a maximal distance of 40 genes between syntenic genes were built with an

in-house perl script, using a single linkage clustering with a Euclidian distance based on the gene index order in each chromosome. These clusters were refined using Synmap (<http://synteny.cnr.berkeley.edu/CoGe/SynMap.pl>) with the BLASTZ algorithm, an average distance expected between syntenic genes of 10, a maximum distance between two matches of 30, a minimum number of aligned gene pairs of 10 and a quota-align ratio of 3 to 3 (Supplementary Text).

For comparative genome analyses, orthologous gene-pairs were identified using predicted proteomes of *M. acuminata*, *O. sativa* (IRGPS/RAP, build 4), *Vitis vinifera* (http://www.genoscope.cns.fr/externe/Download/Projets/Projet_ML/data/12X/annotation/) and *Phoenix dactylifera* (draft sequence version 3, <http://qatar-weill.cornell.edu/research/datepalmGenome/download.html>). Alignments were performed using BLASTP (e value 1×10^{-5}) and retaining best hits. Syntenic clusters of genes were built using a single linkage clustering with a Euclidian distance. Dot-plots were performed using an in-house perl program allowing the painting of paralogous and orthologous gene clusters. Circle diagrams were made with Circos⁴⁸.

To calculate the number of synonymous substitutions per site (Ks), ClustalW⁴⁹ alignments of paralogous and orthologous protein sequences were used to guide nucleic coding sequence alignments with PAL2NAL⁵⁰. Ks values were calculated using the Yang–Nielson method implemented in PAML⁵¹.

Phylogenomic analysis. To infer the timing of genome duplication events relative to speciation events, all annotated *Musa* genes were sorted based on best BLASTP hit into the gene family clusters circumscribed by Jiao *et al.*⁵² and the PlantTribes database⁵³ (<http://fgp.bio.psu.edu/tribedb/>), including sequenced eudicotyledons and monocotyledons, along with transcriptome assemblies for other non-grass monocotyledons (Supplementary Text). Gene family clusters were queried for *Sorghum*¹⁸ and *Musa* orthologues mapping to syntenic blocks, and maximum likelihood gene trees were estimated for these gene families using the GTR+GAMMA model of molecular evolution in RAxML⁵⁴. The estimation of divergence times was performed on maximum likelihood trees based on concatenated MAFFT⁵⁵ alignments for 93 gene families that included only one gene from each of the sequenced genomes (Supplementary Text).

Comparative analysis of gene families. The *Musa* proteome was globally compared with *O. sativa* (RGAP version 6.0), *S. bicolor* (JGI version 1.4), *B. distachyon* (JGI version 1.0), *P. dactylifera* (draft sequence version 3, <http://qatar-weill.cornell.edu/research/datepalmGenome/download.html>) and *A. thaliana* (TAIR version 9) proteomes filtered of transposable elements and alternative splicing. An all-against-all comparison was performed using BLASTP (1×10^{-10}) followed by clustering with OrthoMCL⁵⁶ (inflation 1.5). Analysis of species-specific sets was made with a Fisher's exact test ($P < 0.0001$) on InterPro (version 28) domains. For analyses of specific gene families, the 36,542 *Musa* protein sequences were inserted in the plant proteome clustering of the GreenPhyl database⁵⁷. Transcription factor families were mostly retrieved based on InterPro domains, using the IPR2genomes tool in GreenPhylDB⁵⁷ (Supplementary Text). Kinases and nucleotide-binding site proteins were retrieved using hidden markov models (hmmsearch version 3) to search for corresponding Pfam domains (Supplementary Text).

Identification of CNSs. Pan-grass CNSs conserved between rice, sorghum and *Brachypodium* were prepared using an automated pipeline⁵⁸. The obtained 16,978 CNSs were used to query *Musa* using BLASTN (e value < 0.001) following a manual or a semi-automated procedure depending on CNS size (Supplementary Text and Supplementary Fig. 24). The resulting set of CNSs was extensively analysed using GEvo⁵⁹ (<http://synteny.cnr.berkeley.edu/CoGe/GEvo.pl>) and the MSU Rice Genome Browser⁶⁰ (<http://rice.plantbiology.msu.edu/cgi-bin/gbrowse/rice/>) to remove false positives (Supplementary Table 19). Adding rice and sorghum homeologues, *Brachypodium* and maize orthologues and *Arabidopsis* 'best hit orthologues' to GEvo panels enabled the identification of 18 CNSs conserved deeply throughout the plant kingdom.

Transcriptome sequencing. For RNA-Seq analyses (Supplementary Text), cDNA libraries were sequenced using 76-base length read chemistry in a single-flow cell on the Illumina GA IIx. Reads were mapped against the automatic annotated transcripts with SOAPaligner/Soap2 (2.20, <http://soap.genomics.org.cn/>) and only the unique mapped reads were kept. RNA-seq data were statistically analysed with the R packages baySeq version 1.6.0 (ref. 61) and DESeq version 1.5.6 (ref. 62).

- Bakry, F., Assani, A. & Kerbellec, F. Haploid induction: androgenesis in *Musa balbisiana*. *Fruits* **63**, 45–49 (2008).
- Marie, D. & Brown, S. C. A cytometric exercise in plant DNA histograms, with 2C values for 70 species. *Biol. Cell* **78**, 41–51 (1993).
- Piffanelli, P., Vilarinhos, A., Safar, J., Sabau, X. & Dolezel, J. Construction of bacterial artificial chromosome (BAC) libraries of banana (*Musa acuminata* and *Musa balbisiana*). *Fruits* **63**, 375–379 (2008).
- Aury, J. M. *et al.* High quality draft sequences for prokaryotic genomes using a mix of new sequencing technologies. *BMC Genomics* **9**, 603 (2008).
- Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).

36. Parra, G., Blanco, E. & Guigo, R. GenelD in *Drosophila*. *Genome Res.* **10**, 511–515 (2000).
37. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
38. Salamov, A. A. & Solovyev, V. V. *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res.* **10**, 516–522 (2000).
39. Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res.* **14**, 988–995 (2004).
40. Bairoch, A. *et al.* The Universal Protein Resource (UniProt). *Nucleic Acids Res.* **33**, D154–D159 (2005).
41. Mott, R. EST_GENOME: a program to align spliced DNA sequences to unspliced genomic DNA. *Comput. Appl. Biosci.* **13**, 477–478 (1997).
42. Denoeud, F. *et al.* Annotating genomes with massive-scale RNA sequencing. *Genome Biol.* **9**, R175 (2008).
43. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
44. Argout, X. *et al.* The genome of *Theobroma cacao*. *Nature Genet.* **43**, 101–108 (2010).
45. Flutre, T., Duprat, E., Feuillet, C. & Quesneville, H. Considering transposable element diversification in *de novo* annotation approaches. *PLoS ONE* **6**, e16526 (2011).
46. Ma, J., Devos, K. M. & Bennetzen, J. L. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res.* **14**, 860–869 (2004).
47. Glemet, E. & Codani, J. J. LASSAP, a LARge Scale Sequence compARison Package. *Comput. Appl. Biosci.* **13**, 137–143 (1997).
48. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
49. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
50. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–W612 (2006).
51. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
52. Jiao, Y. *et al.* Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**, 97–100 (2011).
53. Wall, P. K. *et al.* PlantTribes: a gene and gene family resource for comparative genomics in plants. *Nucleic Acids Res.* **36**, D970–D976 (2008).
54. Stamatakis, A. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
55. Katoh, K., Asimenos, G. & Toh, H. Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* **537**, 39–64 (2009).
56. Li, L., Stoeckert, C. J. Jr & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
57. Rouard, M. *et al.* GreenPhylDB v2.0: comparative and functional genomics in plants. *Nucleic Acids Res.* **39**, D1095–D1102 (2011).
58. Woodhouse, M. R. *et al.* Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. *PLoS Biol.* **8**, e1000409 (2010).
59. Lyons, E. & Freeling, M. How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant J.* **53**, 661–673 (2008).
60. Ouyang, S. *et al.* The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res.* **35**, D883–D887 (2007).
61. Hardcastle, T. & Kelly, K. baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics* **11**, 422 (2010).
62. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).

Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation

Sameer A. Sheth^{1*}, Matthew K. Mian^{1*}, Shaun R. Patel^{1,2}, Wael F. Asaad³, Ziv M. Williams¹, Darin D. Dougherty⁴, George Bush⁴ & Emad N. Eskandar¹

The ability to optimize behavioural performance when confronted with continuously evolving environmental demands is a key element of human cognition. The dorsal anterior cingulate cortex (dACC), which lies on the medial surface of the frontal lobes, is important in regulating cognitive control. Hypotheses about its function include guiding reward-based decision making¹, monitoring for conflict between competing responses² and predicting task difficulty³. Precise mechanisms of dACC function remain unknown, however, because of the limited number of human neurophysiological studies. Here we use functional imaging and human single-neuron recordings to show that the firing of individual dACC neurons encodes current and recent cognitive load. We demonstrate that the modulation of current dACC activity by previous activity produces a behavioural adaptation that accelerates reactions to cues of similar difficulty to previous ones, and retards reactions to cues of different difficulty. Furthermore, this conflict adaptation, or Gratton effect^{2,4}, is abolished after surgically targeted ablation of the dACC. Our results demonstrate that the dACC provides a continuously updated prediction of expected cognitive demand to optimize future behavioural responses. In situations with stable cognitive demands, this signal promotes efficiency by hastening responses, but in situations with changing demands it engenders accuracy by delaying responses.

Human cognition is characterized by the ability to parse and evaluate a stream of constantly changing environmental stimuli so as to choose the most appropriate response in evolving conditions. The dACC is thought to be important in regulating cognitive control over goal-directed behaviour. Various theories postulate its involvement in linking reward-related information to action^{1,5,6}, monitoring for conflict between competing responses^{2,7,8} or detecting the likelihood of error commission^{3,9,10}. Despite substantial information from studies using lesions, functional magnetic resonance imaging (fMRI) and event-related potentials, the neurophysiological basis of its regulatory role remains the subject of considerable debate.

We studied dACC function with a combination of fMRI, single-neuronal recordings and observations of behaviour before and after lesion in human subjects undergoing surgical cingulotomy, a procedure in which a precise, stereotactically targeted lesion is created in the dACC. Microelectrode recordings, which are routinely performed during the procedure^{11,12}, allowed us to record from individual dACC neurons. Six subjects participated, and in four of these we also obtained a preoperative fMRI with the same task. In four we recorded behavioural responses using the same task immediately after cingulotomy.

Subjects performed the multi-source interference task¹³, a Stroop-like task in which they viewed a cue consisting of three numbers and had to indicate, by pressing a button, the unique number ('target') that differed from the other two numbers ('distracters') (Fig. 1a). By varying the position of the target and the identity of the distracters, the task

established four distinct trial types (Fig. 1b), which were presented to the subject randomly. These trial types contained three levels (type 0, 1 and 2 trials; Fig. 1b) of cognitive interference, operationally defined here as the tendency of an irrelevant stimulus feature (for example, the position of the target) to impede simultaneous processing of the relevant stimulus feature (for example, the identity of the target).

When high-interference (type 2) trials were compared with non-interference (type 0) trials, there was an increased fMRI signal within the dACC (Fig. 1d), indicating increased neuronal population activity during trials with greater cognitive interference. Other cortical regions known to be involved in this decision-making network, such as the

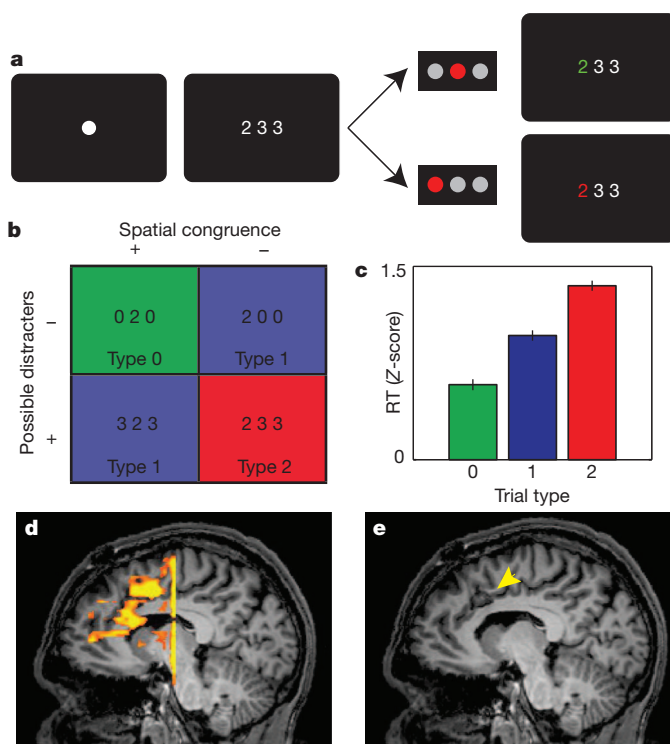


Figure 1 | Behavioural task, fMRI and subject performance. **a**, The multi-source interference task. **b**, The four trial types, based on the presence or absence of spatial congruence between the position of the target and correct button response and on whether the distracters are possible (1, 2, 3) or impossible (0) button choices. **c**, Increase in RTs with increasing cognitive interference in the cue ($P < 10^{-20}$, ANOVA; error bars indicate s.e.m., $n = 1,545$). **d**, Representative example showing greater fMRI activation in the dACC during high-interference trials than in non-interference trials. **e**, A parasagittal slice depicting the lesion (arrowhead), which was also the site of microelectrode recordings.

¹Nayef Al-Rodhan Laboratories, Department of Neurosurgery, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ²Department of Anatomy and Neurobiology, Boston University School of Medicine, Boston, Massachusetts 02118, USA. ³Department of Neurosurgery, Alpert Medical School, Brown University and Rhode Island Hospital, Providence, Rhode Island 02912, USA. ⁴Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.

*These authors contributed equally to this work.

dorsolateral pre-frontal cortex (DLPFC), were similarly activated to a greater degree in the high-interference condition (Supplementary Fig. 1). The spatial distribution and magnitude of these changes were similar to those previously observed in healthy volunteers^{14,15}, suggesting that this function is spared in the dACC in our subject population and is comparable to that in normal subjects. We registered the postoperative MRI (Fig. 1e) with the preoperative fMRI, confirming that the recording and lesion site localized together with the region of fMRI activation.

During intraoperative microelectrode recordings, subjects performed the task accurately, with an error rate of 1.4%. Reaction times (RTs) were modulated by degree of interference in a dose-dependent fashion (Fig. 1c and Supplementary Fig. 2; $P < 10^{-20}$, analysis of variance (ANOVA)). The trial-type-dependent RTs and low error rates were consistent with the tendency to sacrifice speed for accuracy that is often observed in Stroop-like tasks^{8,16}.

We recorded 59 well-isolated single dACC neurons, with an average baseline firing rate of 5.7 ± 0.7 (mean \pm s.e.m.) spikes per second. We identified three distinct subpopulations of neurons on the basis of their maximal task-responsiveness: first, those firing preferentially before the cue ($n = 12$; 20%); second, those firing preferentially after presentation of the cue ($n = 24$; 41%); and third, those firing preferentially after the behavioural choice ($n = 23$; 39%). The largest group, or cue-responsive neurons, showed distinct modulation of firing on the basis of the degree of interference present in the cue ($P = 0.02$, ANOVA; Supplementary Table 1). Paralleling the pattern for RTs, firing rates for type 2 trials were higher than those for type 1 trials, which were higher than those for type 0 trials. This effect during the cue epoch was observable at the level of individual neurons (Fig. 2a, b), as well as at the cue neuron population level (Fig. 2c, d). Inclusion of the entire recorded neuronal population produced similar effects (Supplementary Fig. 3), and using raw rather than normalized rates did not change this result. Neuronal activity within the dACC was thus correlated with the degree of cognitive interference present in the cue.

The trial-type-dependent modulation in firing rate could be a consequence of dACC neuron sensitivity to either the amount of conflict engendered by the cue^{2,17} or the number of potential responses activated by the cue¹⁸ (Supplementary Note 1). To distinguish between these possibilities, we identified trials in which the number of potential

responses remained constant (two), but the amount of conflict varied (one or two types of conflict). Firing rates were significantly higher ($P = 6.4 \times 10^{-3}$, Mann–Whitney test) in higher-conflict trials, indicating that dACC neurons were encoding conflict itself rather than the potential number of responses (Supplementary Fig. 4). RTs for the higher-conflict trials were also significantly higher ($P = 1.5 \times 10^{-4}$, *t*-test), providing behavioural evidence for the increase in perceived conflict. In a two-way ANOVA (with degree of conflict as one variable and the number of possible responses as the other) including all trials, the degree of conflict was a significant independent predictor of firing rate ($P = 5.7 \times 10^{-3}$), whereas the number of possible responses was not ($P = 0.11$).

Current models of dACC function, whether predicated on conflict monitoring^{2,8,17}, reinforcement learning^{3,19} or reward-based decision making^{1,20,21}, require that future dACC activity reflect past experience, but modulation of dACC firing on the basis of recent history has not been demonstrated at the single-neuronal level. To determine whether dACC neuronal firing rates are influenced by previous activity, we separated type 0 and type 2 trials depending on whether they were immediately preceded by a trial containing interference (type 1 or 2) or not (type 0). In both cases, dACC neuronal activity increased more rapidly after the cue when the preceding trial contained interference (Fig. 3a, b). The average magnitude of the cue neuron signal was greater in trials preceded by interference. This finding also held for the entire neuronal population (Supplementary Fig. 5) and was not altered by using raw rather than normalized rates.

The association between previous and current trial activity was maintained across all successive trial pairs. On a trial-by-trial basis including all trial types, previous trial activity during the cue period was significantly correlated with current trial activity ($r = 0.15$, $P = 2.0 \times 10^{-11}$ for cue neurons; $r = 0.12$, $P = 3.3 \times 10^{-16}$ for all neurons; Supplementary Note 2), demonstrating that dACC neurons encode information about both the current task context and the recent past. This neural–neural correlation was not simply an effect of drift in the recordings (Supplementary Note 3).

The behavioural correlate of this neuronal pattern of activity depended on the identity of the current trial. RTs for type 0 trials were positively correlated with previous trial activity ($r = 0.13$,

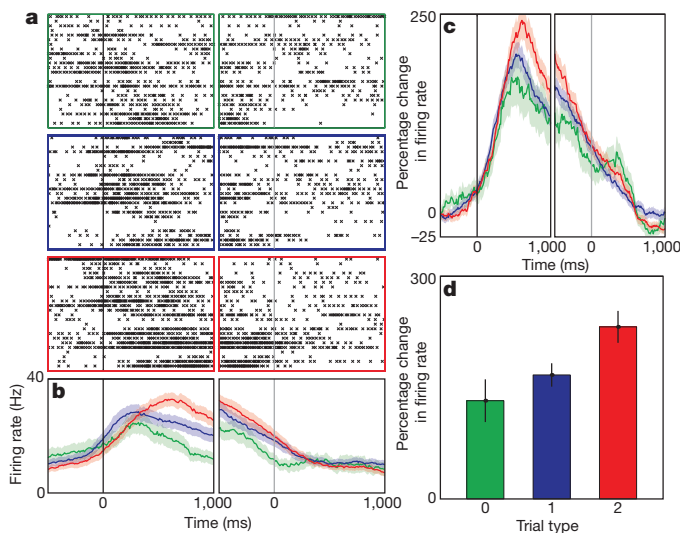


Figure 2 | Individual and population neuronal responses. **a**, Example neuron showing modulation of firing based on cue-related interference. Rasters for type 0 (green), 1 (blue) and 2 (red) trials are shown aligned to the cue (black line) and choice (grey line). **b**, Average firing rates of the same neuron. Error bars (s.e.m., $n = 72$) are depicted with shading. **c**, **d**, Firing rates (**c**) and averaged activity within a window 200 ms wide centred 500 ms after the cue (**d**) for all cue-related neurons. Neuronal firing increased with cognitive interference ($P = 0.02$, ANOVA), correlating with RT.

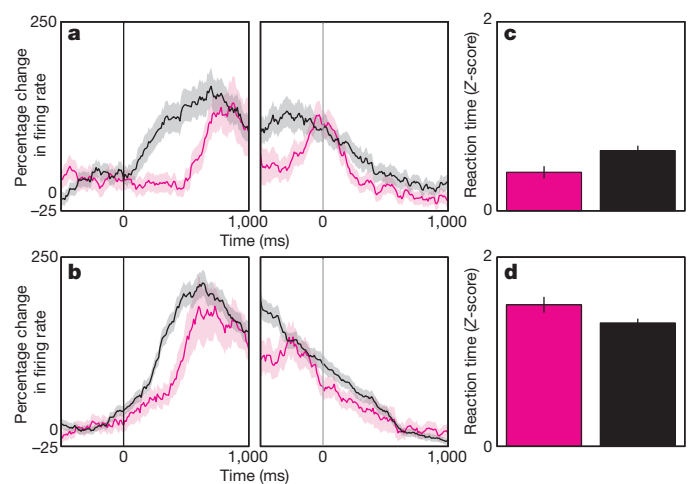


Figure 3 | Effect of previous trial on dACC firing and RT. **a**, Activity was greater for current non-interference trials immediately preceded by interference trials (1,2→0, black) than non-interference trials (0→0, purple). **b**, Similarly, activity was greater for high-interference trials preceded by interference trials (1,2→2, black) than by non-interference trials (0→2, purple). **c**, RTs for non-interference trials were shorter ($P = 0.008$) when preceded by non-interference trials (0→0, purple) than by interference trials (1,2→0, black). **d**, RTs for high-interference trials were shorter ($P = 0.04$) when preceded by interference trials (1,2→2, black) than by non-interference trials (0→2, purple). Shading and error bars represent s.e.m. ($n = 816$).

$P = 9.2 \times 10^{-3}$), meaning that previous elevated activity (consistent with a previous difficult trial) predicted a longer RT on the current non-interference trial. RTs during type 2 trials, however, were negatively correlated with previous trial activity ($r = -0.09$, $P = 8.4 \times 10^{-3}$), meaning that previous elevated activity predicted a shorter RT on the current interference trial. Taken together, these findings predict that RTs for a particular trial type will be shorter when the preceding trial is of the same type, and longer when of a different type (Supplementary Note 4).

The behavioural responses bore out these predictions. RTs during non-interference trials were shorter when preceded by another non-interference (0→0) trial than by an interference (1,2→0) trial (Fig. 3c and Supplementary Fig. 6a). Conversely, RTs during high-interference trials were shorter when preceded by an interference (1,2→2) trial than by a non-interference (0→2) trial (Fig. 3d and Supplementary Fig. 6b). This dependence of dACC neuronal activity on history provides a neurophysiological basis for the current data and for previous observations of behavioural adaptations, known as micro-adjustments, conflict adaptation or the Gratton effect^{2,4,8,17,22,23}.

We analysed post-cingulotomy task performance and thereby captured the acute behavioural manifestations of a precise, stereotyped, reproducible lesion to the previously recorded area (Supplementary Fig. 7). The error rate after cingulotomy was 1.3%, indicating that subjects had not changed in their ability to perform the task. The post-cingulotomy RT distribution was similar to that observed before the lesion (Fig. 4a and Supplementary Fig. 8a): longer RTs were associated with increasing interference ($P < 10^{-12}$, ANOVA). Furthermore, there was no difference in mean RT before and after cingulotomy ($P = 0.76$, *t*-test). Thus, the dACC lesion did not significantly disrupt the subjects' ability to perform the task, nor did it affect the dependence of RT on the cognitive load presented by the current stimulus.

Cingulotomy, however, caused abolition of the history-dependent modulation of RTs. The pre-lesion trial-to-trial adaptations in RT were significantly reduced after cingulotomy for both non-interference trials ($P = 2.2 \times 10^{-8}$, bootstrap test) and high-interference trials ($P = 7.1 \times 10^{-3}$). Consistently, the difference in RT attributable to the previous trial that existed before the lesion (Fig. 3c, d and Supplementary Fig. 6a, b) disappeared after the lesion (Fig. 4b, c and Supplementary Fig. 8b, c); that is, RTs did not depend on the preceding trial type. This effect was observable at the population level and at the level of individual subjects (Supplementary Fig. 9). Thus, although dACC lesions did not globally degrade subject performance, they eliminated the dependence of behavioural responses on recent experience.

These trial-to-trial behavioural adjustments (faster RT when the preceding trial type was of the same type as the current one) are concordant with those reported by others^{2,4,8,17,22,23}. During high-interference trials

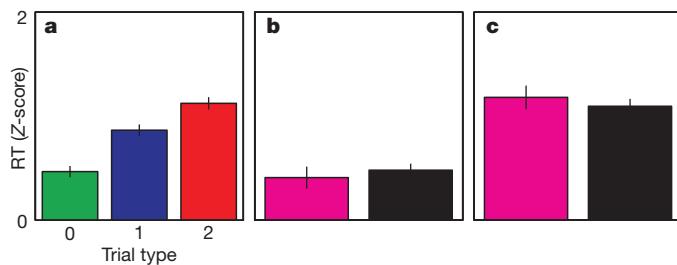


Figure 4 | Abolition of behavioural adaptation after a targeted dACC lesion. RTs were recorded after cingulotomy, in which a stereotactic lesion was created precisely in the region of the dACC from which fMRI signals and microelectrode recordings were obtained. **a**, RTs followed a dose-response pattern by trial type ($P < 10^{-12}$, ANOVA) similar to that obtained before the lesion (Fig. 1c). Error bars represent s.e.m. ($n = 572$). **b**, **c**, However, behavioural adaptations (the influences of previous trial identity on current trial RTs) were abolished for both non-interference trials ($P = 0.54$) (**b**) and high-interference trials ($P = 0.53$) (**c**).

preceded by a high-interference trial, however, we observed increased single-neuron activity in the dACC, whereas others have reported a decreased blood-oxygen-level-dependent fMRI signal^{12,17}. This apparent discrepancy may be explained by the fact that the peak fMRI signal (which occurs 5–7 s after the cue) reflects input synaptic activity evoked by both the appearance of the cue and evaluation of the response (which all occur within the first 1 s after the cue). In contrast, we recorded output spiking activity occurring within 500–750 ms after the cue. These complementary measures may therefore reflect the spatiotemporal dynamics of conflict processing in the dACC.

In this task, increasing interference within the cue could variably be interpreted as representing increasing conflict between competing potential responses⁸, increasing likelihood of error commission³ or increasing energetic cost of decision making²⁰. Consistent with these theories, we find that dACC activity is correlated with these manifestations of cognitive demand. However, although the dACC is modulated by the cognitive load within the context of the current task, its function is not essential for generating the load-dependent behavioural response^{16,24}, because interference-dependent behaviour was not altered after dACC ablation. In contrast, an intact dACC is required for trial-to-trial behavioural adjustments.

Previous studies have proposed that the dACC monitors for conflict between competing responses^{7,8} and drives behaviour towards efficient strategies²⁰. Whereas previous human studies using fMRI^{12,17}, event-related potentials²⁵ and lesions²⁶ have implicated the dACC in this function, single-unit data supporting this theory have been lacking. Moreover, non-human primate single-unit recordings^{18,21} and lesion²⁴ studies have arrived at opposite conclusions and cast doubt on the conflict-monitoring theory²⁷. We reconcile these issues by demonstrating fMRI and single-neuronal conflict signals in the human dACC, and also behavioural adjustments that disappear after a precisely targeted dACC lesion. Our results support the view that the dACC is specifically responsible for providing a continuously updated account of predicted demand on cognitive resources. This account is particularly sensitive to relative shifts in situational complexity from instance to instance, weighted by the recent past. The salient influence of current dACC activity on future neuronal activity and behaviour permits the implementation of behavioural adjustments that optimize performance. In situations in which cognitive demands remain constant, this signal facilitates efficiency by accelerating responses. In situations involving rapidly changing demands, it promotes accuracy by retarding responses.

METHODS SUMMARY

Six human subjects (four males, ages 37.5 ± 5 years) undergoing stereotactic cingulotomy for treatment-refractory obsessive-compulsive disorder provided informed consent and enrolled. The study was approved by our Institutional Review Board. The surgical procedure produces a stereotypical lesion with an average volume of 3.58 ± 1.24 cm³, centred 9 mm lateral to the midline, 18 mm anterior to the anterior commissure, and 30 mm superior to the anterior-commissure-posterior-commissure plane²⁸.

During surgery, subjects performed the multi-source interference task^{13,14}, which was presented on a computer monitor with a customized software package in MATLAB (MathWorks)²⁹. Each trial contained a stimulus consisting of three integers from 0 to 3. One number (the unique 'target') differed from a pair of 'distracter' numbers. Subjects were asked to report, by pressing a button, the identity (rather than the position) of the target (left button for 1, middle for 2, right for 3).

Functional MRI data were analysed with Brain Voyager software (Brain Innovation). Anatomical and functional data were registered and transformed into common Talairach space. A general linear model was constructed by using predictors modelled by convolution with a standard haemodynamic response function. Single-subject repeated-measures ANOVAs were performed on a voxel-wise basis. Multiple comparisons were accounted for by using a cluster constraint with regional false-positive probability $P < 10^{-4}$, requiring a cluster of at least seven contiguous voxels with $P < 0.05$.

For microelectrode recordings, an array of three tungsten microelectrodes (500–1,500 k Ω ; FHC) was attached to a motorized microdrive (Alpha Omega

Engineering). On reaching the cingulate cortex, microelectrodes were held in place and monitored for about 5 min to assess signal stability. Putative neurons were not screened for task responsiveness. Analogue data were amplified, bandpass filtered between 300 Hz and 6 kHz, sampled at 20 kHz and spike-sorted (Offline Sorter).

Full Methods and any associated references are available in the online version of the paper.

Received 9 October 2011; accepted 17 May 2012.

Published online 24 June 2012.

- Williams, Z. M., Bush, G., Rauch, S. L., Cosgrove, G. R. & Eskandar, E. N. Human anterior cingulate neurons and the integration of monetary reward with motor responses. *Nature Neurosci.* **7**, 1370–1375 (2004).
- Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S. & Cohen, J. D. Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature* **402**, 179–181 (1999).
- Brown, J. W. & Braver, T. S. Learned predictions of error likelihood in the anterior cingulate cortex. *Science* **307**, 1118–1121 (2005).
- Gratton, G., Coles, M. G. & Donchin, E. Optimizing the use of information: strategic control of activation of responses. *J. Exp. Psychol. Gen.* **121**, 480–506 (1992).
- Hayden, B. Y. & Platt, M. L. Neurons in anterior cingulate cortex multiplex information about reward and action. *J. Neurosci.* **30**, 3339–3346 (2010).
- Narayanan, N. S. & Laubach, M. Neuronal correlates of post-error slowing in the rat dorsomedial prefrontal cortex. *J. Neurophysiol.* **100**, 520–525 (2008).
- Botvinick, M. M., Cohen, J. D. & Carter, C. S. Conflict monitoring and anterior cingulate cortex: an update. *Trends Cogn. Sci.* **8**, 539–546 (2004).
- Carter, C. S. & van Veen, V. Anterior cingulate cortex and conflict detection: an update of theory and data. *Cogn. Affect. Behav. Neurosci.* **7**, 367–379 (2007).
- Carter, C. S. *et al.* Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science* **280**, 747–749 (1998).
- Nieuwenhuis, S., Schweizer, T. S., Mars, R. B., Botvinick, M. M. & Hajcak, G. Error-likelihood prediction in the medial frontal cortex: a critical evaluation. *Cereb. Cortex* **17**, 1570–1581 (2007).
- Davis, K. D., Hutchison, W. D., Lozano, A. M., Tasker, R. R. & Dostrovsky, J. O. Human anterior cingulate cortex neurons modulated by attention-demanding tasks. *J. Neurophysiol.* **83**, 3575–3577 (2000).
- Davis, K. D. *et al.* Human anterior cingulate cortex neurons encode cognitive and emotional demands. *J. Neurosci.* **25**, 8402–8406 (2005).
- Bush, G. & Shin, L. M. The Multi-Source Interference Task: an fMRI task that reliably activates the cingulo-frontal-parietal cognitive/attention network. *Nature Protocols* **1**, 308–313 (2006).
- Bush, G., Shin, L. M., Holmes, J., Rosen, B. R. & Vogt, B. A. The Multi-Source Interference Task: validation study with fMRI in individual subjects. *Mol. Psychiatry* **8**, 60–70 (2003).
- Bush, G. *et al.* Functional magnetic resonance imaging of methylphenidate and placebo in attention-deficit/hyperactivity disorder during the multi-source interference task. *Arch. Gen. Psychiatry* **65**, 102–114 (2008).
- Fellows, L. K. & Farah, M. J. Is anterior cingulate cortex necessary for cognitive control? *Brain* **128**, 788–796 (2005).
- Kerns, J. G. *et al.* Anterior cingulate conflict monitoring and adjustments in control. *Science* **303**, 1023–1026 (2004).
- Nakamura, K., Roesch, M. R. & Olson, C. R. Neuronal activity in macaque SEF and ACC during performance of tasks involving conflict. *J. Neurophysiol.* **93**, 884–908 (2005).
- Holroyd, C. B. & Coles, M. G. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol. Rev.* **109**, 679–709 (2002).
- Botvinick, M. M. Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cogn. Affect. Behav. Neurosci.* **7**, 356–366 (2007).
- Ito, S., Stuphorn, V., Brown, J. W. & Schall, J. D. Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science* **302**, 120–122 (2003).
- Mayr, U., Awh, E. & Laurey, P. Conflict adaptation effects in the absence of executive control. *Nature Neurosci.* **6**, 450–452 (2003).
- Ridderinkhof, K. R. Micro- and macro-adjustments of task set: activation and suppression in conflict tasks. *Psychol. Res.* **66**, 312–323 (2002).
- Mansouri, F. A., Buckley, M. J. & Tanaka, K. Mnemonic function of the dorsolateral prefrontal cortex in conflict-induced behavioral adjustment. *Science* **318**, 987–990 (2007).
- Gehring, W. J. & Fencsik, D. E. Functions of the medial frontal cortex in the processing of conflict and errors. *J. Neurosci.* **21**, 9430–9437 (2001).
- di Pellegrino, G., Ciaramelli, E. & Ladavas, E. The regulation of cognitive control following rostral anterior cingulate cortex lesion in humans. *J. Cogn. Neurosci.* **19**, 275–286 (2007).
- Cole, M. W., Yeung, N., Freiwald, W. A. & Botvinick, M. Cingulate cortex: diverging data from humans and monkeys. *Trends Neurosci.* **32**, 566–574 (2009).
- Rauch, S. L. *et al.* Volume reduction in the caudate nucleus following stereotactic placement of lesions in the anterior cingulate cortex in humans: a morphometric magnetic resonance imaging study. *J. Neurosurg.* **93**, 1019–1025 (2000).
- Asaad, W. F. & Eskandar, E. N. A flexible software tool for temporally-precise behavioral control in Matlab. *J. Neurosci. Methods* **174**, 245–258 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from the National Science Foundation (IOB 0645886), the National Institutes of Health (NEI 1R01EY017658-01A1, NIDA 1R01NS063249, NIMH Conte Award MH086400 and R25 NS065743), the Klingenstein Foundation, the Howard Hughes Medical Institute, the Sackler Scholar Programme in Psychobiology, the Centers for Disease Control (5 R01 DP000339), the Benson-Henry Institute at Massachusetts General Hospital for Mind-Body Medicine, the David Judah Fund, the McIngvale Fund, and the Center for Functional Neuroimaging Technologies (P41RR14075).

Author Contributions E.N.E., W.F.A., Z.M.W. and D.D.D. designed the study. G.B. administered and interpreted the fMRI scans. E.N.E. performed the surgical procedures, and S.A.S., M.K.M., S.R.P. and W.F.A. obtained the neuronal recordings. S.A.S. and M.K.M. analysed the data and wrote the manuscript. All authors edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to E.N.E. (eeskandar@partners.org).

METHODS

Subjects. We enrolled six study subjects (four male, ages 37.5 ± 5 years (mean \pm s.e.m.)) undergoing stereotactic cingulotomy for treatment-refractory obsessive-compulsive disorder. Evaluation for surgical candidacy was conducted by a multidisciplinary team consisting of psychiatrists, neurologists and neurosurgeons. Subjects enrolled voluntarily, providing informed consent under a protocol approved by the Massachusetts General Hospital Institutional Review Board. The surgical procedure produces a stereotypical lesion with an average volume of 3.58 ± 1.24 cm³ (mean \pm s.d.), centred 9 mm lateral to the midline, 18 mm anterior to the anterior commissure, and 30 mm superior to the anterior-commissure-posterior-commissure plane²⁸. Subject participation was in no way related to clinical decision-making regarding their candidacy for surgery.

Behavioural task. During the microelectrode recording portion of surgery, subjects performed the multi-source interference task (MSIT) (Fig. 1a, b)^{13,14}. The task was presented on a computer monitor using a customized software package in MATLAB (MathWorks)²⁹. Each trial contained a stimulus consisting of three integers from 0 to 3. One number (the unique 'target') differed from a pair of 'distracter' numbers (for example 100 or 323). Subjects were asked to report, by pressing a button, the identity (rather than the position) of the target (left button for 1, middle for 2, right for 3).

Functional MRI. Functional MRI was performed before surgery by using a 3.0-T scanner (Allegra, Siemens AG) and head coil. The MSIT was presented on a screen visible by means of a tilted mirror, and controlled with MacStim 2.6 software (WhiteAnt Occasional Publishing). Scans were acquired with the following specifications: 15 coronal sections, 64×64 matrix, 3.125 mm² in-plane resolution, 5 mm thickness with 0 mm skip, 30 ms echo time, 1,500 ms repetition time, 90° flip angle, 20 cm² field of view. During fMRI, only types 0 and 2 trials (no interference and both types of interference; see Fig. 1b) were used. The task was run in a block design. Each block consisted of 24 trials of the same type. One run consisted of eight alternating blocks, with an additional five visual fixation blocks interspersed. Data were analysed with Brain Voyager software (Brain Innovation). Anatomical and functional data were registered and transformed into common Talairach space. A general linear model was constructed by using predictors modelled by convolution with a standard haemodynamic response function. Single-subject repeated-measures ANOVAs were performed on a voxel-wise basis. Multiple comparisons were accounted for by using a cluster constraint with regional false-positive probability $P < 10^{-4}$. This constraint required a cluster of at least seven contiguous voxels with $P < 0.05$.

Surgical procedure. The surgical procedure was performed with standard stereotactic techniques. A Cosman-Roberts-Wells (Integra) stereotactic frame was affixed to the patient under local anaesthesia, and a high-resolution MRI was obtained. The target for the left posterior lesion (2 cm posterior to the most anterior point of the frontal horn of the lateral ventricle, 0.7 cm lateral to midline, and 0.5 cm superior to the corpus callosum) was programmed into a neuro-navigation computer (StealthStation, Medtronic) and the stereotactic frame was then set appropriately. The patient was positioned semi-recumbent, the surgical area was prepared, and sterile drapes were applied. Local anaesthetic was infiltrated, a coronal skin incision was performed, and bilateral burr holes were drilled 1.5 cm lateral to the midline and 10.0 cm posterior to the nasion. A computerized microelectrode drive controlled by a neurophysiology system (Alpha Omega) was affixed to the frame. After dural opening, microelectrodes were lowered using the computerized drive in increments of 0.01 mm. The position of the tip of the electrodes was also monitored in real time using the stereotactic neuronavigation system. After microelectrode recordings, a thermocoagulation electrode with a 10-mm exposed tip (Cosman Medical) was lowered to the target. Lesions were performed by heating the electrode to 85 °C for 60 s. Two more pairs of lesions were then created, each 7 mm anterior and 2 mm inferior to the previous lesion.

Microelectrode recordings. For microelectrode recordings, an array of three tungsten microelectrodes (500–1,500 k Ω ; FHC) was attached to a motorized microdrive (Alpha Omega Engineering). As per routine surgical protocol, recordings were obtained from the left hemisphere. On reaching the cingulate cortex, microelectrodes were held in place and monitored for about 5 min to assess signal stability. Putative neurons were not screened for task responsiveness. Analogue data were amplified, bandpass filtered between 300 Hz and 6 kHz, sampled at 20 kHz (Alpha Omega Engineering) and spike-sorted (Offline Sorter, Plexon).

Post-lesion behaviour. Immediately after creation of the cingulate lesions, subjects again performed the MSIT. These task sessions were identical in all respects to the pre-lesion sessions except that we collected only behavioural data. Four of the six patients participated in these post-lesion sessions.

RT data analysis. RT was defined as the interval between the onset of the cue and the subject's button-press. To allow for comparisons between subjects, we normalized subject RTs relative to their individual distributions. Normalized RTs were defined as the number of standard deviations (Z-scores) above or below the bottom tenth centile of a subject's RT distribution. The choice of each subject's reference point for the normalization (bottom tenth centile versus median or mean) is arbitrary and does not affect the result, because it simply represents a rigid translation of the normalized values. We chose the bottom tenth centile so that values were positive, to facilitate visual comparisons.

We also computed a speed of target selection (STS) as described in ref. 24. We inverted RT to obtain $STS = 1/RT$. We then divided by the mean STS across all trials in the behavioural session. STS was thus normalized to 1, which represented an intermediate or 'average' response speed. Faster responses (for example, on low-conflict trials) are reflected as relative increases in STS (that is, $STS > 1$), whereas slower responses (for example, on high-conflict trials) are attended by decrements in the STS (that is, $STS < 1$). Use of the normalized STS, rather than raw RTs, allowed us to view data from different subjects on a common scale and to minimize between-subject differences when pooling data to calculate composite statistics.

Single-unit data analysis. Single units were isolated by first building a histogram of peak heights from the raw voltage tracings on each channel. We applied a minimum threshold of three standard deviations to exclude background noise. Action potentials were sorted by using waveform principal component analysis. Spike clusters of putative neurons were required to separate clearly from any channel noise, to demonstrate a voltage waveform consistent with that of a cortical neuron and to have at least 99% of action potentials separated by an inter-spike interval of 1 ms or more.

We recorded an average of 1.1 neurons per microelectrode. When a single channel captured more than one neuron cluster, we required a clear distinction between the two clusters to include either one as a single unit ($P < 0.01$, multivariate ANOVA across the first two principal components). In addition, we required putative dACC neurons to fire at an average rate of at least 1.0 spikes per second, to be stably active for at least 25 task trials, and to not demonstrate significant drift over the duration of the recording. We excluded single units not meeting these criteria. We did not use any multi-unit activity.

We classified neurons into three mutually exclusive categories on the basis of the timing of their peak firing rates with respect to task events. Neurons peaking in activity before stimulus presentation, during the stimulus period and after the button-push were classified as 'pre-cue', 'cue' and 'feedback' neurons, respectively.

To facilitate comparisons between neurons with different firing rates, we normalized (divided) neuronal activity by the average neuronal firing rate during a 500-ms window preceding the appearance of the fixation point. Population firing rates were computed by averaging these normalized neuronal responses with a 250-ms moving boxcar window.

HVEM signalling at mucosal barriers provides host defence against pathogenic bacteria

Jr-Wen Shui¹, Alexandre Larange¹, Gisen Kim¹, Jose Luis Vela¹, Sonja Zahner¹, Hilde Cheroutre¹ & Mitchell Kronenberg¹

The herpes virus entry mediator (HVEM), a member of the tumour-necrosis factor receptor family, has diverse functions, augmenting or inhibiting the immune response¹. HVEM was recently reported as a colitis risk locus in patients², and in a mouse model of colitis we demonstrated an anti-inflammatory role for HVEM³, but its mechanism of action in the mucosal immune system was unknown. Here we report an important role for epithelial HVEM in innate mucosal defence against pathogenic bacteria. HVEM enhances immune responses by NF- κ B-inducing kinase-dependent Stat3 activation, which promotes the epithelial expression of genes important for immunity. During intestinal *Citrobacter rodentium* infection^{4–6}, a mouse model for enteropathogenic *Escherichia coli* infection, *Hvem*^{−/−} mice showed decreased Stat3 activation, impaired responses in the colon, higher bacterial burdens and increased mortality. We identified the immunoglobulin superfamily molecule CD160 (refs 7 and 8), expressed predominantly by innate-like intraepithelial lymphocytes, as the ligand engaging epithelial HVEM for host protection. Likewise, in pulmonary *Streptococcus pneumoniae* infection⁹, HVEM is also required for host defence. Our results pinpoint HVEM as an important orchestrator of mucosal immunity, integrating signals from innate lymphocytes to induce optimal epithelial Stat3 activation, which indicates that targeting HVEM with agonists could improve host defence.

Because HVEM functions in the colonic mucosa^{2,3,10}, we explored its role in colonic epithelial cells, a cell type whose barrier function is critical for preventing colitis pathogenesis and enhancing mucosal

defence¹¹. In addition to leukocytes that carry the CD45 antigen, HVEM is expressed by epithelial cells in the colon and lung, as well as by mouse epithelial cell lines CMT-93 (colon) and LA-4 (lung) (Supplementary Fig. 1). When stimulated *in vitro* by known HVEM ligands, either a fusion protein containing the B- and T-lymphocyte attenuator (BTLA-Ig) or recombinant CD160, CMT-93 cells, colon fragments in culture or primary colonic epithelial cells could be induced to produce innate immune mediators essential for host protection (Fig. 1a and Supplementary Fig. 2). These included genes encoding anti-microbial proteins, such as Reg3 β /3 γ , β -defensin 3 and S100A9, proinflammatory cytokines IL-6, IL-1 β and tumour-necrosis factor (TNF), as well as chemokines. As epithelial cells produce these mediators, our data suggest that it was HVEM signalling in epithelium that regulated innate immune responses in the colon fragments. Interestingly, the innate immune mediators that were increased in epithelial cells are also induced by IL-22R signalling, which acts through Stat3 (refs 12 and 13). Although, like other TNF receptors, HVEM signals through TRAF proteins to activate NF- κ B¹⁴, we found that HVEM engagement by its ligands also induced Stat3 phosphorylation in epithelial cells and colon fragment cultures (Fig. 1b and Supplementary Fig. 3). Interestingly, HVEM engagement did not induce immediate Stat3 activation in leukocytes, and although HVEM signalling has been reported to promote NF- κ B-inducing kinase (NIK)-Stat3-dependent Th17 cell differentiation *in vitro*¹⁵, this may require extra and/or more prolonged stimulation. By contrast, the rapid HVEM-mediated Stat3 activation in epithelial cells emphasizes

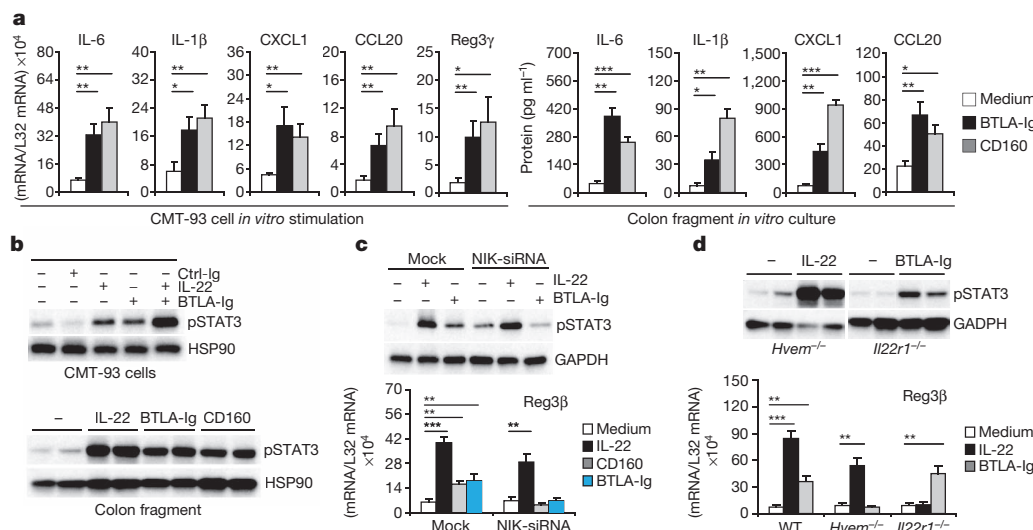


Figure 1 | HVEM signalling regulates epithelial immune function by inducing NIK-dependent Stat3 activation. **a**, HVEM signalling triggered by soluble BTLA-Ig or CD160 induced innate responses in CMT-93 epithelial cells analysed by real-time PCR (left), or in colon fragments analysed by enzyme-linked immunosorbent assay (ELISA) (right). **b**, HVEM signalling induced Stat3 phosphorylation in CMT-93 cells and colon fragments (each lane represents one

mouse), determined by western blot. **c**, NIK dependence of HVEM-induced epithelial Stat3 activation and gene expression analysed by NIK short interfering RNA knockdown. **d**, HVEM regulates epithelial Stat3 activation and gene expression independent of IL-22R signalling. Error bars (s.e.m.) are indicated. **P* < 0.05, ***P* < 0.01, ****P* < 0.001 (two-tailed unpaired *t*-tests). Results were representative of at least two independent experiments.

¹Division of Developmental Immunology, La Jolla Institute for Allergy and Immunology, 9420 Athena Circle, La Jolla, California 92037, USA.

the direct connection of HVEM signals to Stat3 phosphorylation. We further determined that HVEM-induced Stat3 activation was NIK-dependent (Fig. 1c and Supplementary Fig. 4). As Stat3 activity is associated with epithelial responses¹³, our results indicate that NIK, upstream of Stat3, links epithelial HVEM signalling to mucosal innate immunity and host defence.

IL-22 is essential for host defence because it induces epithelial production of cytokines, chemokines and anti-microbial peptides, and promotes epithelial recovery from mucosal injury^{12,16–21}. Although IL-22 signalling also induces epithelial Stat3 activation and gene expression, we found signalling by IL-22R did not require NIK (Fig. 1c). This suggests that the HVEM–Stat3 pathway is independent of the IL-22–Stat3 pathway. Indeed, we found that IL-22 induced Stat3 activation and epithelial gene expression in the absence of HVEM signalling and vice versa (Fig. 1d). However, HVEM and IL-22 signalling exhibited additive effects on Stat3 activation (Fig. 1b) and gene expression (Supplementary Fig. 2). As such, we reasoned that a fully activated Stat3, cooperatively regulated by IL-22R and HVEM, probably is important for intestinal epithelial responses and host defence.

We tested this hypothesis using *C. rodentium* infection, a mouse model for acute attaching/effacing enteropathogenic *E. coli* infection in humans⁴. We found that *Hvem*^{−/−} mice had reduced survival, higher bacterial burdens in colons and faeces and increased colonic epithelial permeability after infection and bacterial dissemination (Fig. 2a–c and Supplementary Fig. 5). The compromised epithelial integrity and responses after infection probably contributed to bacterial translocation and lethality. As a consequence of the poorly controlled bacterial burdens, *Hvem*^{−/−} mice suffered more severe colonic hyperplasia (at day 14) and pathology (at day 14 and 21) after infection (Fig. 2c and Supplementary Fig. 5). Analysis of infected *Hvem*^{−/−} *Rag*^{−/−} mice also showed an impaired production of epithelial mediators and increased pathology compared with *Rag*^{−/−} mice (Supplementary Fig. 6), indicating a role for HVEM expression independent of adaptive immunity. Consistent with a role for epithelial HVEM, using bone-marrow chimaeric mice, we found that HVEM expression by radiation-resistant cells in the recipients was required for host protection (Fig. 2d).

Because HVEM binds to the TNF superfamily member LIGHT, and immunoglobulin (Ig) superfamily members BTLA and CD160 (ref. 1), we performed experiments to identify which is the important ligand(s) for HVEM in the intestine. *Light*^{−/−}, *Btla*^{−/−} and *Light*^{−/−} *Btla*^{−/−} double knockout mice were able to clear bacteria similarly to wild-type mice (Supplementary Fig. 7), suggesting that neither HVEM–LIGHT nor HVEM–BTLA signalling is required for host defence against *C. rodentium*. By contrast, mice injected with an anti-CD160 blocking but non-depleting antibody showed impaired host defence (Fig. 3a and Supplementary Figs 8 and 9), indicating that CD160 is essential. We performed colon fragment cultures, *in vitro* infected with *C. rodentium*, which targets epithelial cells^{5,22}, to explore further the

ligand necessary for HVEM engagement. We found *Hvem*^{−/−}, *Il-22r1*^{−/−} or anti-CD160 treated wild-type colons, but not *Light*^{−/−} or *Btla*^{−/−} colons, showed reduced expression of host defence mediators (Supplementary Fig. 10 and data not shown). CD160 also interacts with some major histocompatibility complex (MHC) class I molecules⁷, but CD160 blockade in *Hvem*^{−/−} colons did not cause an alteration in the innate immune response or survival in infected mice (Supplementary Fig. 10).

To determine the *in vivo* function of HVEM–CD160 signalling further, we analysed the caecum at day 2 after infection, because *C. rodentium* is known to colonize this part of the intestine first⁴. We found *Hvem*^{−/−} and *Il-22r1*^{−/−} mice, as well as anti-CD160 injected mice, had impaired caecal expression of host defence genes (Fig. 3b). Cytokine and chemokine protein production was also reduced in the colons of infected *Hvem*^{−/−} mice (Supplementary Fig. 11). Furthermore, caecal Stat3 activation was significantly reduced in *Hvem*^{−/−} and *Il-22r1*^{−/−} mice, as well as in anti-CD160 injected mice (Fig. 3c and Supplementary Fig. 12). These results indicate that by regulating epithelial Stat3 activation, both the CD160–HVEM and IL-22–IL-22R signalling pathways are essential for epithelial innate function and mucosal host defence. Supporting this, we observed that mice with an epithelial cell-specific IL-22R1 deletion (*Vil-Cre; Il-22r1*^{flx/flx}) were also susceptible to *C. rodentium* infection, similar to *Hvem*^{−/−} mice (Supplementary Fig. 10d).

We determined if the two pathways of Stat3 activation in epithelial cells influence one another. Similar to the action of several other cytokines on their receptors, IL-22 upregulated expression of its own receptor (Fig. 3d), specifically the IL-22R1 subunit, but not the IL-10R2 chain shared with several other cytokines. Interestingly, expression of IL-22R1 was also increased after HVEM engagement in CMT-93 cells and colon fragments. Furthermore, early after bacterial infection, colonic IL-22R1 expression was reduced in *Hvem*^{−/−} and anti-CD160-injected mice (Fig. 3d and Supplementary Fig. 13). These results indicate that HVEM, in addition to activating Stat3 independently of the IL-22 receptor, could cross-regulate IL-22R signalling and sensitize epithelium by promoting IL-22R1 expression. This would render epithelial cells more responsive to bacterial infection. Consistent with this model for the regulation of receptor expression is the fact that the human and mouse *Il22r1* promoters contain both Stat and NF-κB binding sites^{23,24}, which could be the targets of Stat3 and NF-κB activated by HVEM²⁵.

Epithelial Reg3β/γ was shown to be a major target of Stat3 signalling during mucosal inflammation¹³ and was important for host protection against *C. rodentium*¹². As HVEM induced epithelial Reg3 expression, and there was a dramatic decrease in Stat3 activation and Reg3β/γ expression in *Hvem*^{−/−} colons during infection, it is likely that colonic Reg3 expression, mediated by HVEM–Stat3, contributed to host protection. Indeed, recombinant Reg3γ protein administered to *Hvem*^{−/−} mice protected them from lethal infection and significantly reduced bacterial burdens in faeces (Supplementary Fig. 14). These results

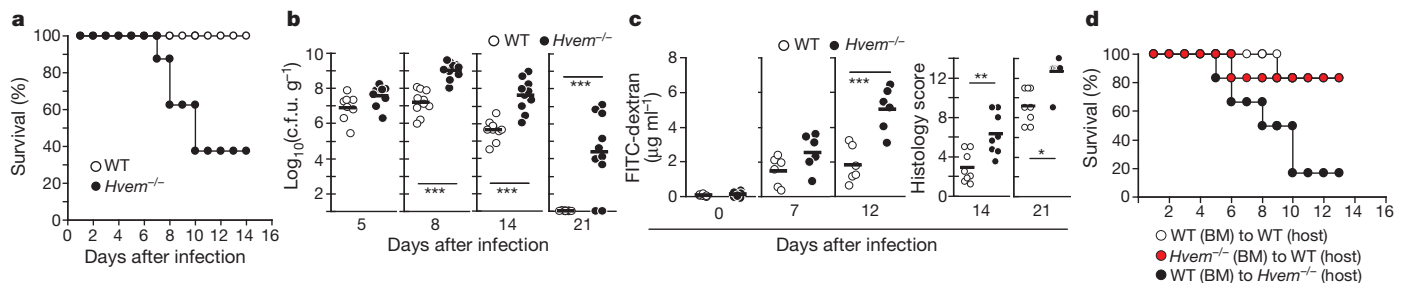


Figure 2 | HVEM is required for host defence against intestinal *C. rodentium* infection. Groups of mice ($n = 6$ –12) were infected with *C. rodentium* by oral gavage. **a**, Survival curves for wild-type and *Hvem*^{−/−} mice after infection. **b**, Bacterial burdens in colons after infection. **c**, Epithelial barrier permeability (left), determined by injection of fluorescein isothiocyanate

(FITC)-dextran and measurement of fluorescence in the blood, and severe colonic pathology in *Hvem*^{−/−} mice (right). **d**, HVEM expression by non-bone-marrow-derived cells is required for host protection ($n = 6$ in each group of chimaeras). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (two-tailed unpaired *t*-tests). Results were representative of at least two independent experiments.

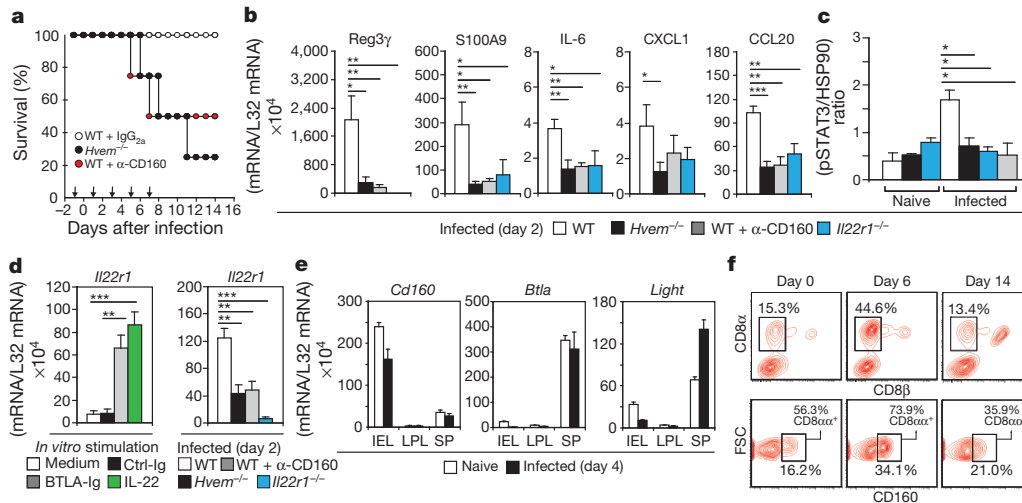


Figure 3 | CD160 on IELs provides a ligand for HVEM. **a**, Survival curves of wild-type (WT), *Hvem*^{-/-} and anti-CD160-injected mice (*n* = 4) after *C. rodentium* infection. Arrows indicate the times of control IgG or anti-CD160 injection. **b**, **c**, Analysis of gene expression (by real-time PCR) and Stat3 phosphorylation (by western blot) in caecal tissues collected from indicated mice at day 2 after infection. **d**, Analysis of *Il22r1* mRNA in stimulated or

infected colon fragments. **e**, Analysis of *Cd160*, *Btla* and *Light* mRNA in colonic IELs, LPLs and splenocytes, isolated from naive or infected mice. **f**, Flow cytometry analysis of colonic CD45⁺ IELs, isolated from naive or infected mice. **P* < 0.05, ***P* < 0.01, ****P* < 0.001 (two-tailed unpaired *t*-tests). Results were representative of two independent experiments.

indicate that the HVEM–Stat3–Reg3 pathway in the epithelium provides host protection during intestinal infection. Both BTLA and CD160 could trigger HVEM signalling and induce epithelial Stat3 activation and production of host defence mediators *in vitro*; however, it is still uncertain why only CD160 was essential for host protection *in vivo*. When expression levels of the three HVEM ligands in colonic intraepithelial lymphocytes (IELs), lamina propria lymphocytes (LPLs) and splenocytes were compared, we found that CD160 was almost exclusively expressed by IELs (Fig. 3e). There were very low messenger RNA (mRNA) levels of BTLA and LIGHT in IELs, providing an explanation why these were not functional ligands for HVEM after *C. rodentium* infection. A more detailed analysis of the IEL and LPL compartments during bacterial infection revealed that the CD8α⁺CD8β⁺CD160⁺ IEL population was rapidly increased at the early stage of infection (Fig. 3f and Supplementary Fig. 15). These CD160⁺ IELs with the CD8αα homodimer expression consisted of both T-cell receptor (TCR)αβ and TCRγδ T subsets, but the TCRγδ IELs became more prevalent after infection (Supplementary Fig. 15). Furthermore, CD160 was also expressed by TCR-negative IELs in *Rag*^{-/-} mice, which presumably provide a physiological ligand for epithelial HVEM in the absence of adaptive immunity (Supplementary Fig. 16). Together, these results provide evidence that CD160

expressed by several IEL subsets, particularly the innate-like CD8αα-expressing cells, represents the only ligand available for engaging epithelial HVEM in the intestine. Furthermore, the increase in CD8αα⁺CD160⁺ IELs during early infection is consistent with the hypothesis that the HVEM–CD160 interaction at the mucosal surface enhances signalling in epithelial cells and promotes their innate response to acute bacterial infection. It remains possible, however, that the contribution of the HVEM–CD160 signal is regulated in other ways, for example by the removal of an inhibitor in one of the interacting cell types.

To determine if HVEM mediates epithelial host defence at other mucosal sites, we examined the immune response of *Hvem*^{-/-} mice after lung infection with *S. pneumoniae*, a Gram-positive extracellular bacteria that is the most common cause of community-acquired pneumonia^{26,27}. We found *Hvem*^{-/-} mice were highly susceptible to *S. pneumoniae*, with an impaired early host response leading to reduced bacterial elimination in the lung (Fig. 4a). Similar to the role of HVEM in colonic epithelium, the reduced innate immune response in the lung epithelium early after infection probably contributed to the impaired host defence in *Hvem*^{-/-} mice (Fig. 4b and Supplementary Fig. 17). Stat3 signalling in the alveolar epithelium is essential for host defence to limit bacteria-induced pneumonia^{26,28,29}. Similar to the

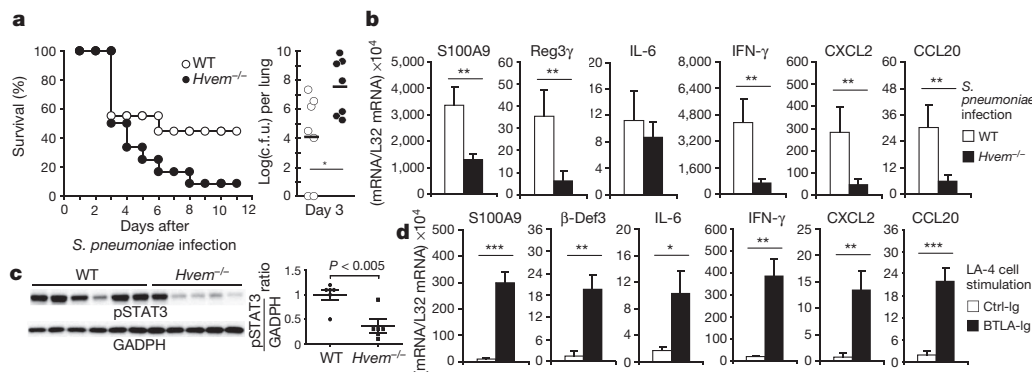


Figure 4 | HVEM is required for lung epithelial immunity and host defence against *S. pneumoniae*. Groups of mice (*n* = 9–12) were retropharyngeally infected with *S. pneumoniae* and analysed at the indicated time points. **a**, Survival curves of mice and *S. pneumoniae* c.f.u. in the lung at day 3 after infection. **b**, mRNA expression by real-time PCR in the lung at day 3 after

infection. **c**, Analysis of Stat3 phosphorylation in the lung of *S. pneumoniae*-infected mice (each lane represents an individual mouse) at day 2 after infection. **d**, mRNA analysis by real-time PCR in LA-4 lung epithelial cells, stimulated with control-Ig or BTLA-Ig. **P* < 0.05, ***P* < 0.01, ****P* < 0.001 (two-tailed unpaired *t*-tests). Results were representative of two independent experiments.

intestinal mucosae, HVEM stimulation induced Stat3 phosphorylation in the lung and Stat3 activation was significantly decreased in the lung of *Hvem*^{-/-} mice at the early stage of *S. pneumoniae* infection (Fig. 4c and Supplementary Fig. 18). This indicates HVEM–Stat3 signalling regulates epithelial responses in the lung and mediates host defence against pulmonary bacteria. LA-4 cells, a lung epithelial cell line, also express HVEM (Supplementary Fig. 1c). When these cells were stimulated *in vitro* to trigger HVEM signalling, the expression of several epithelial host defence genes was significantly induced (Fig. 4d), indicating HVEM directly regulates lung epithelial immune function.

HVEM is an unusual TNF receptor family member in binding TNF as well as Ig superfamily molecules, and in serving as a ligand for inhibitory receptors such as BTLA, as well as being a signalling receptor itself. We have revealed that HVEM, acting as a signalling receptor on epithelial cells, leads to NIK-dependent Stat3 activation, thereby inducing the expression of genes important for host defence in the intestine and lung. Furthermore, we demonstrated that in the intestine CD160 on CD8 α ⁺ IELs is the functional ligand for triggering epithelial HVEM for host protection. Human CD160 expression on IELs and epithelial HVEM expression have been described^{8,30}, but it is not known if this interaction has a role in humans. HVEM signalling also promotes epithelial IL-22R1 expression, which renders epithelial cells more responsive to IL-22, a major innate cytokine critical for host defence. Our results therefore show that HVEM is a crucial regulator of epithelial innate responses, in part by mediating lymphocyte–epithelial communication, and by cooperating with IL-22R signalling to induce optimal epithelial Stat3 activation for host defence.

METHODS SUMMARY

Mice. C57BL/6 and *Rag*^{-/-} mice were purchased from the Jackson Laboratory and bred in-house for all experiments. *Hvem*^{-/-}, *Hvem*^{-/-}*Rag*^{-/-}, *Btla*^{-/-}, *Light*^{-/-} and *Light*^{-/-}*Btla*^{-/-} mice have been bred and described previously³. *Il-22r1* flox (knockout first) mice on the C57BL/6 background were obtained from The European Conditional Mouse Mutagenesis Program (EUCOMM) and were bred to *Vil-Cre* mice for conditional IL-22R1 deletion in epithelium (*Vil-Cre*;*Il-22r1*^{flox/flox}). All gene-deficient mice were backcrossed for at least eight generations onto the C57BL/6 background. Mice were maintained at the La Jolla Institute for Allergy and Immunology under specific pathogen-free conditions, and sentinels from the mouse colony tested negative by PCR detection of *Helicobacter* spp. Animal care and experimentation including infection protocols were consistent with National Institutes of Health guidelines and were approved by the Institutional Animal Care and Use Committee at the La Jolla Institute for Allergy and Immunology.

Bacterial infection. A wild-type *C. rodentium* strain DBS100 rendered resistant to chloramphenicol and a clinical strain of *S. pneumoniae* (serotype 3, URF918) were obtained and used in all infection studies^{6,9}. For *C. rodentium*, bacteria were grown overnight in Luria-Bertani broth with shaking at 37 °C for 15–16 h. For *S. pneumoniae*, bacteria were grown to mid-log phase for 6–7 h from a frozen stock in Todd-Hewitt broth. Bacterial cultures were adjusted with PBS for proper concentration and individual titres were determined after each experiment by serial dilution. Mice were infected with 0.5×10^9 to 3.5×10^9 colony-forming units (c.f.u.) (*C. rodentium*, by oral gavage) or 1×10^6 to 5×10^6 c.f.u. (*S. pneumoniae*, by retropneumal instillation), and were killed at the indicated time points after infection. For c.f.u. assays, spleen, lung, liver, mesenteric lymph node, faecal pellets or colon were weighed, homogenized, serially diluted and plated in chloramphenicol-containing MacConkey (for *C. rodentium*) or Columbia sheep blood (for *S. pneumoniae*) agar plates.

Full Methods and any associated references are available in the online version of the paper.

Received 7 October 2011; accepted 18 May 2012.

Published online 15 July; corrected online 8 August 2012 (see full-text HTML for details).

1. Murphy, T. L. *et al.* Slow down and survive: enigmatic immunoregulation by BTLA and HVEM. *Annu. Rev. Immunol.* **28**, 389–411 (2010).
2. Anderson, C. A. *et al.* Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nature Genet.* **43**, 246–252 (2011).

3. Steinberg, M. W. *et al.* A crucial role for HVEM and BTLA in preventing intestinal inflammation. *J. Exp. Med.* **205**, 1463–1476 (2008).
4. Mundy, R. *et al.* *Citrobacter rodentium* of mice and man. *Cell. Microbiol.* **7**, 1697–1706 (2005).
5. Ma, C. *et al.* *Citrobacter rodentium* infection causes both mitochondrial dysfunction and intestinal epithelial barrier disruption *in vivo*: role of mitochondrial associated protein (Map). *Cell. Microbiol.* **8**, 1669–1686 (2006).
6. LeBlanc, P. M. *et al.* Caspase-12 modulates NOD signaling and regulates antimicrobial peptide production and mucosal immunity. *Cell Host Microbe* **3**, 146–157 (2008).
7. Maeda, M. *et al.* Murine CD160, Ig-like receptor on NK cells and NKT cells, recognizes classical and nonclassical MHC class I and regulates NK cell activation. *J. Immunol.* **175**, 4426–4432 (2005).
8. Anumanthan, A. *et al.* Cloning of BY55, a novel Ig superfamily member expressed on NK cells, CTL, and intestinal intraepithelial lymphocytes. *J. Immunol.* **161**, 2780–2790 (1998).
9. Nakamatsu, M. *et al.* Role of interferon-gamma in V α 14⁺ natural killer T cell-mediated host defense against *Streptococcus pneumoniae* infection in murine lungs. *Microbes Infect.* **9**, 364–374 (2007).
10. Schaer, C. *et al.* HVEM signalling promotes colitis. *PLoS ONE* **6**, e18495 (2011).
11. Artis, D. *et al.* Epithelial-cell recognition of commensal bacteria and maintenance of immune homeostasis in the gut. *Nature Rev. Immunol.* **8**, 411–420 (2008).
12. Zheng, Y. *et al.* Interleukin-22 mediates early host defense against attaching and effacing bacterial pathogens. *Nature Med.* **14**, 282–289 (2008).
13. Pickert, G. *et al.* STAT3 links IL-22 signaling in intestinal epithelial cells to mucosal wound healing. *J. Exp. Med.* **206**, 1465–1472 (2009).
14. Marsters, S. A. *et al.* Herpesvirus entry mediator, a member of the tumor necrosis factor receptor (TNFR) family, interacts with members of the TNFR-associated factor family and activates the transcription factors NF- κ B and AP-1. *J. Biol. Chem.* **272**, 14029–14032 (1997).
15. Jin, W. *et al.* Regulation of Th17 cell differentiation and EAE induction by MAP3K NIK. *Blood* **113**, 6603–6610 (2009).
16. Liang, S. C. *et al.* Interleukin (IL)-22 and IL-17 are coexpressed by Th17 cells and cooperatively enhance expression of antimicrobial peptides. *J. Exp. Med.* **203**, 2271–2279 (2006).
17. Sugimoto, K. *et al.* IL-22 ameliorates intestinal inflammation in a mouse model of ulcerative colitis. *J. Clin. Invest.* **118**, 534–544 (2008).
18. Wolk, K. *et al.* Biology of interleukin-22. *Semin. Immunopathol.* **32**, 17–31 (2010).
19. Zenewicz, L. A. *et al.* Innate and adaptive interleukin-22 protects mice from inflammatory bowel disease. *Immunity* **29**, 947–957 (2008).
20. Wolk, K. *et al.* IL-22 increases the innate immunity of tissues. *Immunity* **21**, 241–254 (2004).
21. Brand, S. *et al.* IL-22 is increased in active Crohn's disease and promotes proinflammatory gene expression and intestinal epithelial cell migration. *Am. J. Physiol. Gastrointest. Liver Physiol.* **290**, G827–G838 (2006).
22. Torchinsky, M. B. *et al.* Innate immune recognition of infected apoptotic cells directs Th(H)17 cell differentiation. *Nature* **458**, 78–82 (2009).
23. Tachiiri, A. *et al.* Genomic structure and inducible expression of the IL-22 receptor alpha chain in mice. *Genes Immun.* **4**, 153–159 (2003).
24. Gelebart, P. *et al.* Interleukin 22 signaling promotes cell growth in mantle cell lymphoma. *Transl. Oncol.* **4**, 9–19 (2011).
25. Cheung, T. C. *et al.* Unconventional ligand activation of herpesvirus entry mediator signals cell survival. *Proc. Natl Acad. Sci. USA* **106**, 6244–6249 (2009).
26. Quinton, L. J. *et al.* Functions and regulation of NF- κ B RelA during pneumococcal pneumonia. *J. Immunol.* **178**, 1896–1903 (2007).
27. Mizgerd, J. P. *et al.* Animal models of human pneumonia. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **294**, L387–L398 (2008).
28. Quinton, L. J. *et al.* NF- κ B and STAT3 signaling hubs for lung innate immunity. *Cell Tissue Res.* **343**, 153–165 (2011).
29. Quinton, L. J. *et al.* Alveolar epithelial STAT3, IL-6 family cytokines, and host defense during *Escherichia coli* pneumonia. *Am. J. Respir. Cell Mol. Biol.* **38**, 699–706 (2008).
30. Rooney, I. A. *et al.* The lymphotoxin-beta receptor is necessary and sufficient for LIGHT-mediated apoptosis of tumor cells. *J. Biol. Chem.* **275**, 14307–14315 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from the National Institutes of Health (R01-AI061516 to M.K.; P01 DK46763 to M.K.; F32-DK082249 to J.-W.S.; F32-AI083029 to J.L.V.), La Jolla Institute for Allergy and Immunology and the Center for Infectious Disease (LIAI-JAN-2011-CID to J.-W.S.). We thank W. Quyang for providing Reg3 γ -Ig fusion protein, and O. Turovskaya for performing histology staining and pathological scoring. We also thank C. Benedict for providing *NIK*^{sh/aly} mice, K. Pfeffer for providing *Hvem*^{-/-} and *Light*^{-/-} mice and K. Murphy for *Btla*^{-/-} mice. This is manuscript number 1347 from the La Jolla Institute for Allergy and Immunology.

Author Contributions J.-W.S., A.L., G.K., J.L.V. and S.Z. designed and performed the experiments. H.C. and M.K. contributed to experimental design. J.-W.S. and M.K. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.K. (mitch@liai.org).

METHODS

Animal manipulations. C57BL/6 and *Rag*^{-/-} mice were purchased from the Jackson Laboratory and bred in-house for all experiments. *Hvem*^{-/-}, *Hvem*^{-/-}*Rag*^{-/-}, *Btla*^{-/-}, *Light*^{-/-} and *Light*^{-/-}*Btla*^{-/-} mice have been bred and described previously³. *Il-22r1* flox (knockout first) mice on the C57BL/6 background were obtained from The European Conditional Mouse Mutagenesis Program (EUCOMM) and were bred to *Vil-Cre* mice for conditional IL-22R1 deletion in epithelium (*Vil-Cre;Il-22r1^{flox/flox}*). All gene-deficient mice were backcrossed for at least eight generations onto the C57BL/6 background. Mice were used at 8–12 weeks of age. Whenever possible, control and gene knockout mice were housed in the same cage to minimize the effect of housing conditions on experimental variation. For tissue or cell analyses, tissues were collected and used for histological analysis, intraepithelial and lamina propria cell preparation, total protein or RNA isolation. For the rescue experiment, Reg3γ-Ig fusion protein (150 µg per mouse per time point, provided by Genentech) was injected intraperitoneally into mice on days 1, 3, 5, 7 and 9 after infection as previously described¹². For the blocking experiment, anti-mouse CD160 or IgG_{2a} isotype control antibodies (100 µg per mouse per time point) was injected intraperitoneally into mice on days -1, 1, 3, 5, and 7 after infection.

Bacterial infection. A wild-type *C. rodentium* strain DBS100 rendered resistant to chloramphenicol and a clinical strain of *S. pneumoniae* (serotype 3, URF918) were obtained and used in all infection studies^{6,9}. For *C. rodentium*, bacteria were grown overnight in Luria-Bertani broth with shaking at 37 °C for 15–16 h. For *S. pneumoniae*, bacteria were grown to mid-log phase for 6–7 h from a frozen stock in Todd-Hewitt broth. Bacterial cultures were adjusted with PBS for proper concentration and individual titres were determined after each experiment by serial dilution. Mice were infected with 0.5×10^9 to 3.5×10^9 c.f.u. (*C. rodentium*, by oral gavage) or 1×10^6 to 5×10^6 c.f.u. (*S. pneumoniae*, by retropharyngeal instillation), and were killed at the indicated time points after infection. For c.f.u. assays, spleen, lung, liver, mesenteric lymph node, faecal pellets or colon were weighed, homogenized, serially diluted and plated in chloramphenicol-containing MacConkey (for *C. rodentium*) or Columbia sheep blood (for *S. pneumoniae*) agar plates.

Epithelial cell lines and reagents. Mouse colonic (CMT-93, number CCL-223) and lung (LA-4, number CCL-196) epithelial cell lines were obtained from American Type Culture Collection and were maintained in DMEM (CMT-93) or F12-K (LA-4) medium (Invitrogen), supplemented with 10–15% FBS, penicillin and streptomycin. Cells were cultured in six-well (for protein analysis) or 12-well (for quantitative real-time PCR analysis) plates and were serum starved for 24 h before *in vitro* stimulation. Recombinant mouse IL-22, CD160, LIGHT, BTLA Fc chimaeric fusion protein (BTLA-Ig) and mouse IgG_{2a} Fc (control-Ig) were purchased from R&D Systems and used for *in vitro* stimulation experiments at the indicated concentrations. Anti-mouse CD160 (CNX46-3) and IgG_{2a} isotype control antibodies (both from eBioscience) were used for *in vivo* blocking experiments. A short interfering RNA Kit for NIK knockdown was obtained from Qiagen (GS53859). The short interfering RNA sequences for NIK (*Map3k14*) knockdown were *Map3k14_1* (CTGGGT CAGCTCATAAAGCAA), *Map3k14_2* (CCCTTGAAAGGAGAATATAA), *Map3k14_3* (TAGCATTAAGTCACTGTGAA) and *Map3k14_4* (CAGGAAGATAGTCTCCACTA).

Ex vivo infection of colons. Distal colons were freshly isolated from mice, cleaned and cut into several small pieces (~3 mm). Colon fragments were then transferred to six-well plate containing 3 ml of DMEM supplemented with 10% FBS and antibiotics. Live *C. rodentium* were added to each well (10^7 – 10^8 c.f.u. ml⁻¹) and incubated for 20–24 h. After incubation, *ex vivo* infected colons were collected, briefly washed, and total RNA was extracted using an RNeasy Kit (Qiagen).

Preparation of colonic epithelial cells. Primary epithelial cells were isolated as described previously³¹. Briefly, colons were cut open longitudinally, and epithelial cells were separated by shaking the small colon pieces in HBSS containing 5% FBS, 5 mM EDTA and 1 mM DTT for 15 min. The remaining tissue was discarded and epithelial cells in the supernatant were spun down at 150g for 5 min. The cell pellets were re-suspended in 40% Percoll solution and spun down again. The epithelial cells at the top layer were collected. These cells were stained with epithelial cell specific markers anti-cytokeratin-18 (C-04, Abcam) and anti-EpCAM (G8.8, eBioscience) to estimate purity (>95%).

Preparation of IELs and LPLs. Small and large intestines were collected from mice. Peyer's patches were carefully removed and tissues were cut open longitudinally, briefly washed, and cut into 1.5 cm pieces. The tissue pieces were incubated in 30 ml of HBSS (5% FBS, 10 mM HEPES and 1 mM DTT) in a shaker at 250 r.p.m., 37 °C, for 30 min. After incubation, the cell suspension was intensively vortexed and filtered through a metal mesh. The tissue debris was saved for LPL preparation and the flow-through cell suspension, which constitutes the epithelial cell content and IELs, was spun down at 230g for 5 min. The cell pellets were then re-suspended in 40% Percoll solution and overlaid above a 70% Percoll solution.

The gradient was spun at 800g for 25 min and IELs at the interface after spin were collected. For LPL preparation, the tissue debris was incubated in 20 ml of HBSS (5% FBS, 5 mM EDTA) in a shaker at 250 r.p.m., 37 °C, for 20 min to further remove epithelial cells. After that, tissues were further cut into 1 mm pieces, placed in 20 ml pre-warmed digestion solution containing 1.5 mg ml⁻¹ collagenase type VIII (Sigma) and incubated at 37 °C for 20 min with rotation. After incubation, digested tissues were filtered through a metal mesh. The flow-through cell suspension was spun down and the re-suspended cells were further purified by Percoll gradient centrifugation. LPLs were collected at the interface, washed once and re-suspended in complete RPMI-1640 medium. The cells were used immediately for cell counting and staining.

Generation of bone-marrow chimaeras. Recipient mice (wild-type or *Hvem*^{-/-} in C57BL/6, Ly5.2 background) were lethally irradiated with 12 Gy that were applied in two doses of 6 Gy separated by a 3 h interval, as previously described³. After irradiation, hosts were transplanted with 7×10^6 total bone-marrow cells isolated from wild-type congenic C57BL/6 or *Hvem*^{-/-} mice on the Ly5.1 background to generate three groups ($n = 6$ in each group) of chimaeric animals as follows: WT(Ly5.1) bone-marrow to WT(Ly5.2) host, WT(Ly5.1) bone-marrow to *Hvem*^{-/-}(Ly5.2) host, and *Hvem*^{-/-}(Ly5.1) bone-marrow to WT(Ly5.2) host. After transplant, mice were maintained with antibiotic treatment for the first week and an additional 11 weeks without antibiotic treatment. Mice were checked at 8 weeks after transplant for reconstitution by surface staining of peripheral blood cells for Ly5.1⁺ expression (>95% in lymphocytes in all cases).

FITC-dextran permeability assay. *In vivo* permeability assays to assess intestinal barrier function were performed using FITC-labelled dextran (Sigma) as described previously³². Briefly, mice were infected with *C. rodentium*. At day 7 or day 12 after infection, mice were gavaged with FITC-dextran (60 mg per 100 g body weight) and serum was collected retro-orbitally 7–8 h later. Blood cells in serum samples were removed and the fluorescence intensity of FITC-dextran was measured. FITC-dextran concentrations were determined from standard curves generated by serial dilution.

Western blotting and ELISA. Cell lines or colon fragments, were *in vitro* stimulated as indicated; to reduce background from the cell lines they were serum starved for 24 h. Cells or tissues were directly lysed in RIPA buffer supplemented with a protease and phosphatase inhibitor cocktail (Thermo Scientific). Protein lysates were analysed by SDS-polyacrylamide gel electrophoresis (Bio-Rad) and membranes were blotted with anti-Stat3 and anti-phosphorylated Stat3 antibodies (Cell Signaling). For ELISA, supernatants from colon fragment cultures were collected and analysed by R&D Quantikine ELISA Kits according to the manufacturer's instructions.

Flow cytometry and antibodies. Flow cytometry analysis was on an LSRII instrument (BD Biosciences) and we analysed data using Flowjo software (Tree Star). The following mouse antibodies were purchased from BD Pharmingen or eBiosciences: CD45 (30-F11), TCRβ (H57-597), TCRγδ (eBio-GL3), CD8α (53-6.7), CD8β (eBioH35-17.2), BTLA (6F7), pStat3 (clone 4/P-STAT3), IL-1β (NJTEN3), TNF (MP6-XT22) and CD160 (CNX46-3). Anti-mouse HVEM (HMHV-1B18) antibody was from Biolegend. Anti-mouse IL-22R1 (clone 496514), anti-CCL20 (clone 114906) and anti-CXCL1-Biotin antibodies were purchased from R&D Systems.

Histology analysis. Histopathological analysis of colon samples was performed on formalin-fixed tissue after haematoxylin and eosin staining. Five different histological categories (oedema, inflammation severity, internal bleeding in the gut tissue, crypt damage, percentage of involvement) were scored with a maximum score of 2 or 3 in each category. Maximal total score is 14 for severity of colitis.

Real-time PCR analysis. Total RNA was extracted from infected tissues using an RNeasy Kit (Qiagen), according to the manufacturer's instructions. cDNA synthesis was performed with an iScript cDNA Synthesis Kit (Bio-rad). Quantitative real-time PCR reactions were performed with the SYBR Green I Master Kit and LightCycler 480 system (Roche). All mRNA levels shown in figures were normalized to the housekeeping gene *L32* and a multiplier of 10^4 was indicated when compared with *L32* levels. The primer sets used were described previously^{12,33}.

Statistics. All data were analysed using GraphPad Prism 5 software and were shown as mean and the standard error of the mean (s.e.m.). Two-tailed unpaired Student's *t*-tests were used to analyse the results. Differences indicated and considered significant when $P < 0.05$.

31. Zaki, M. H. *et al.* The NLRP3 inflammasome protects against loss of epithelial integrity and mortality during experimental colitis. *Immunity* **32**, 379–391 (2010).
32. Lebeis, S. L. *et al.* Interleukin-1 receptor signaling protects mice from lethal intestinal damage caused by the attaching and effacing pathogen *Citrobacter rodentium*. *Infect. Immun.* **77**, 604–614 (2009).
33. Ishigame, H. *et al.* Differential roles of interleukin-17A and -17F in host defense against mucocutaneous bacterial infection and allergic responses. *Immunity* **30**, 108–119 (2009).

(Supplementary Fig. 2)⁸, the *ced-3* shed cells were reactive to the phosphatidylserine-binding protein MFG-E8 and to TUNEL staining (Fig. 1h, i). Also, *ced-3*-mutant shed cells showed chromatin condensation (darkly staining nuclear material) and separation of the nuclear envelope double membrane in transmission electron micrographs (Fig. 1f and Supplementary Fig. 3). These apoptotic features were present in *ced-5*-mutant floaters (Fig. 1g and Supplementary Fig. 3), although the cytoplasm of *ced-5*-mutant floaters were more compact. We conclude that the shed cells of embryos lacking caspase activity are in many respects cytologically and morphologically apoptotic, indicating that caspases are dispensable for many cellular changes that occur during apoptosis.

The somatic cell lineage of *C. elegans* is essentially invariant^{10,11}, allowing the precise identification of cell origins and fates. To determine the cellular identities of *ced-3*-mutant shed cells, we recorded time-lapse videos of developing *ced-3*-mutant embryos and traced the lineages of extruded cells in reverse (Fig. 2a, b and Supplementary Movie 1). We identified seven different cells eliminated by shedding from *ced-3*-mutant embryos (Fig. 2c), all of which are cells that normally die during wild-type embryogenesis. This finding is consistent with our observation that *ced-3*-mutant shed cells expressed *egl-1* (Fig. 2d), the transcription of which initiates programmed cell death¹², and with a previous report in which ABalpapap (a cell fated to die) was observed detaching from a *ced-3*-mutant embryo⁵. The cells that can be shed are among the first to die in the wild-type embryo: 14 cells die within the first 300 min of development, and 7 of the 8 identified shed cells are among this group of 14 cells¹¹. Thus, specific cells fated to die early in embryogenesis can be eliminated by either canonical caspase-dependent apoptosis or by caspase-independent shedding.

To identify factors required for cell shedding, we tested genes involved in different cell-death processes and found that the generation of *ced-3*-mutant shed cells did not require genes that mediate germline apoptosis, cell-corpse engulfment, necrosis, autophagy or *lin-24*-mediated cell death (Supplementary Tables 1 and 3). We therefore developed a screening strategy based on the hypothesis that ABalpapap (one of the shed cells we identified; Fig. 2a, c) might

survive and adopt a fate associated with a lineally related cell in animals doubly defective in the canonical cell-death pathway and cell shedding. The sister cell of ABalpapap generates RMEV (a GABAergic (γ -aminobutyric acid-containing) neuron) and the excretory cell (Fig. 3a), which functions in osmoregulation¹³. We generated a *pig-12::cNLS::gfp* transgene to express green fluorescent protein (GFP) specifically in the excretory cell and observed that wild-type and *ced-3*-mutant animals contained a single GFP-positive cell (Fig. 3b, d). Using this reporter, we screened mutagenized *ced-3*-mutant animals and identified double mutants with a two-excretory-cell (Tex) phenotype. Two such Tex isolates, *n5433* and *n5437*, are alleles of the gene *pig-1* (Supplementary Fig. 4). Single mutants defective in a null allele of *pig-1* (*gm344*) had one excretory cell (Fig. 3d), whereas 89% of *pig-1* (*gm344*) *ced-3* (*n3692*) double mutants contained two GFP-positive nuclei that resembled the large nucleus of the excretory cell (Fig. 3c, d and Supplementary Fig. 4). Double mutants with *pig-1* and *ced-4*, *ced-9* (gain-of-function) or *egl-1* mutations were similar to *pig-1* *ced-3* mutant animals (Supplementary Table 4). Inactivation of *pig-1* by RNA interference (RNAi) treatment phenocopied *gm344*, *n5433* and *n5437*, confirming that loss of *pig-1* function caused the Tex phenotype in these mutants (Supplementary Table 7).

The LIN-3 epidermal growth factor ligand is expressed embryonically by the excretory cell¹⁴, and *pig-1* *ced-3* mutant embryos contained an extra cell that expressed *lin-3* (Supplementary Fig. 5a). Furthermore, the heads of *pig-1* *ced-3* mutant animals contained large cysts (Fig. 3e) similar to those of mutants with defective excretory cell function¹³. Thus, *pig-1* *ced-3* mutants generated an ectopic excretory cell, albeit one that either was defective in osmoregulation or interfered with the function of the endogenous excretory cell.

To address whether the ectopic excretory cell of *pig-1* *ced-3* mutants is derived from the un-shed ABalpapap cell, we examined directly the fate of ABalpapap in *pig-1* and *pig-1* *ced-3* mutant embryos (Supplementary Movies 2 and 3). In *pig-1*-mutant embryos (and as in wild type), ABalpapap became a highly refractile cell corpse within 45 min of its generation (Fig. 3f; data not shown). By contrast, ABalpapap survived and divided approximately 115 min after it was generated in *pig-1* *ced-3* mutant embryos (Fig. 3g). In the three *pig-1* *ced-3* mutant embryos we examined, neither ABalpapap nor its descendants detached from the embryo, suggesting that an ABalpapap descendant gives rise to the ectopic excretory cell in *pig-1* *ced-3* mutant larvae. As reported previously¹⁵, *pig-1* *ced-3* animals also contain ectopic RME-like neurons (Supplementary Fig. 5b and Supplementary Table 5), suggesting that when ABalpapap survives it generates both an ectopic RME-like and an ectopic excretory-like cell.

pig-1 inactivation by mutation or RNAi treatment reduced the number of shed cells in *ced-3*-, *ced-4*- or *ced-9* (gain-of-function)-mutant embryos by nearly 75% ($P < 5.0 \times 10^{-7}$ for each pair-wise comparison, Student's *t*-test; Fig. 3h; data not shown), demonstrating that *pig-1* is generally required for the generation of shed cells. Given this observation, the effects of *pig-1* on cell shedding are in several ways comparable to the effects of *ced-3* on programmed cell death: (1) *ced-3* affects most programmed cell deaths and *pig-1* similarly affects most shed cells, indicating that *ced-3* and *pig-1* act generally to drive programmed cell death and generate shed cells, respectively; (2) like programmed cell deaths, extruded cells share morphologic and genetic properties and can be viewed as expressing a specific cell fate; and (3), like mutations in *ced-3*, mutations in *pig-1* cause cells that should die to express the fates of cells that normally survive.

pig-1 encodes a homologue of MELK, an AMPK-related serine-threonine kinase required cell autonomously for the asymmetric cell divisions of many *C. elegans* neuroblasts^{15,16}. Mammalian AMPK-related kinases control metabolism and cell polarity¹⁷ and are activated through phosphorylation of a conserved threonine within their T-loop domains by the kinase LKB1 (also known as STK11) (ref. 18) and its complex partners STRAD α and MO25 α (also known as CAB39) (refs 19, 20). The PIG-1 T-loop threonine (T169) was necessary for a

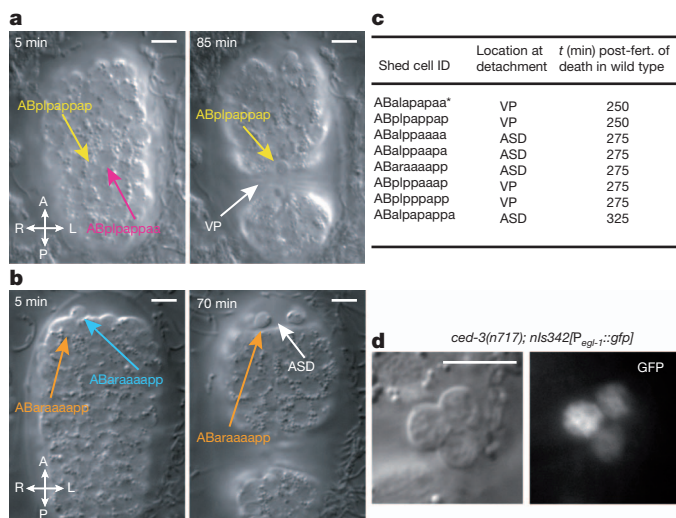


Figure 2 | Cells that are shed from *ced-3*-mutant embryos are normally fated to die early during wild-type embryogenesis. **a, b**, DIC micrographs of *ced-3* (*n3692*) embryos showing ABalpapap (**a**) and ABalpapap (**b**) 5 min after generation and shortly after shedding from the embryo (85 and 70 min later, respectively). ASD, anterior sensory depression; VP, ventral pocket. **c**, Cells that can be shed from *ced-3* (*n3692*) embryos, their locations when extruded and the timings of their deaths in wild-type embryos¹¹. Asterisk denotes data reported in ref. 5. **d**, DIC and fluorescence micrographs of shed cells from a *ced-3* (*n717*) embryo containing the *nls342[P_{egl-1}::gfp]* transgene, which expresses GFP from the *egl-1* promoter. Scale bars, 10 μ m.

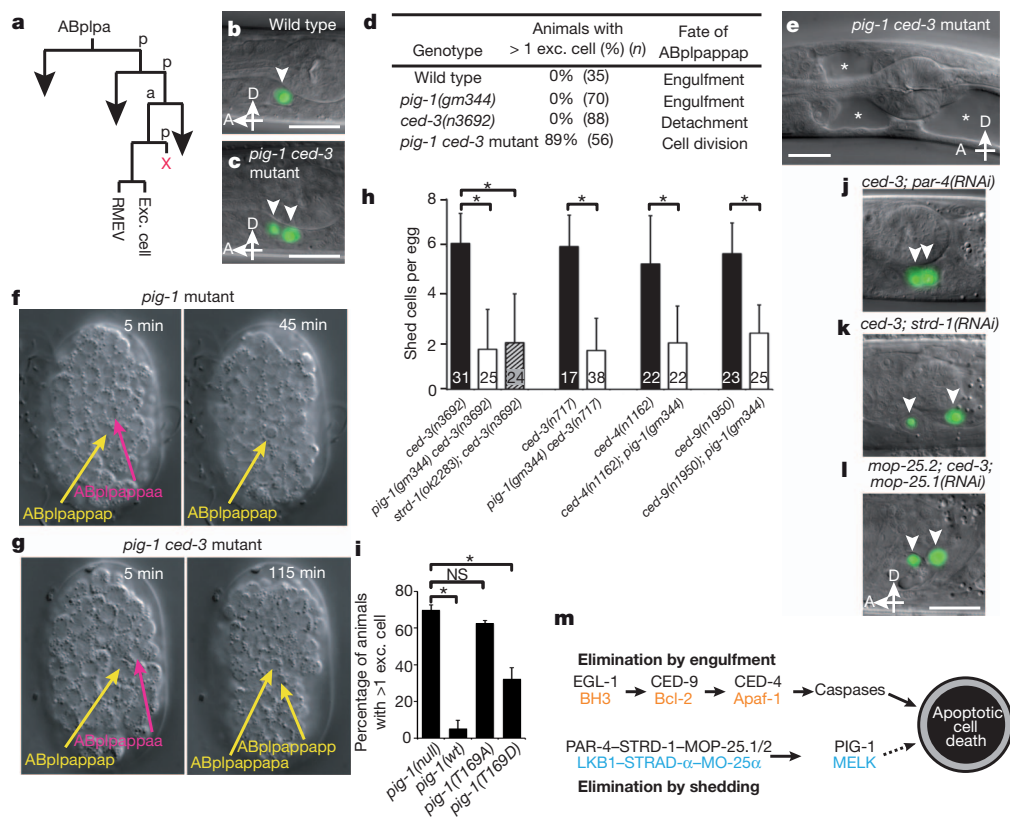


Figure 3 | The LKB1 homologue PAR-4 and the AMPK-related kinase PIG-1 are required for cell shedding from *ced-3* embryos. **a**, The sub-lineage that produces the shed cell ABp1pappap, which is the lineal aunt of the neuron RMEV and the excretory cell (exc. cell). **b**, **c**, Merged DIC and fluorescence micrographs of wild-type (**b**) and *pig-1(gm344) ced-3(n3692)* (**c**) larvae containing the transgene *nIs434[P_{pgp-12::gfp}]*, which expresses GFP in the excretory cell. Arrowheads indicate excretory and ectopic excretory-like cells. **d**, Percentage of L3 larvae with ectopic excretory cells. All genotypes contained *nIs434[P_{pgp-12::gfp}]*. **e**, The head of a larval *pig-1(gm344) ced-3(n3692)* animal containing large cysts (asterisks). **f**, **g**, The fate of the cell ABp1pappap in *pig-1(gm344)* and *pig-1(gm344) ced-3(n3692)* embryos. **f**, ABp1pappap in a *pig-1* mutant embryo shown 5 min after its generation and shortly after it underwent programmed cell death 45 min later. **g**, ABp1pappap in a *pig-1 ced-3* mutant embryo shown 5 min after its generation and immediately after it divided 115 min later. **h**, *pig-1* and *strd-1* are required for cell shedding. Mutation of *pig-1* and of *strd-1* reduced the number of shed cells in *ced-3*, *ced-4* or *ced-9* (gain-of-

pig-1 transgene to rescue the ectopic excretory cell defect of *pig-1 ced-3* mutants, whereas changing T169 to a phosphomimetic aspartic acid bypassed this requirement (Fig. 3i), indicating that PIG-1 is probably activated by phosphorylation of its T-loop. We therefore tested the *C. elegans* homologues of LKB1, STRAD and MO25 (*par-4* and paralogues *mop-25.1*, *mop-25.2* and *mop-25.3*, respectively) for roles in the elimination of ABp1pappap. *par-4* or *strd-1* inactivation caused the Tex phenotype in *ced-3* mutants (Fig. 3j, k and Supplementary Table 6). Furthermore, the *ok2283* deletion allele of *strd-1* reduced the number of shed cells in *ced-3*-mutant embryos by 67% ($P = 1.9 \times 10^{-10}$, Student's *t*-test; Fig. 3h). The inactivation of both *mop-25.1* and *mop-25.2* was necessary to cause the Tex defect in *ced-3* mutants (Fig. 3l and Supplementary Table 6), indicating redundant function. We conclude that both PIG-1 and the PAR-4-STRD-1-MOP-25 complex are required for cell shedding in *C. elegans*.

Our genetic analyses suggest that PIG-1 and PAR-4 function in the same pathway. The deletion mutation *ok2283*, a putative null allele of *strd-1* (Supplementary Table 8), failed to enhance any of the *pig-1* or *pig-1 ced-3* defects (Fig. 4h, i and Supplementary Fig. 6 and Supplementary Tables 8 and 9). Although, *par-4* and *strd-1* can act independently²¹, this result suggests that PIG-1 and PAR-4 and its binding

function (gf) eggs. Error bars denote s.d. $P < 5 \times 10^{-7}$, Student's *t*-test. **i**, The T-loop threonine (T169) of PIG-1 is required for the elimination of ABp1pappap. Average percentage of larvae with ectopic excretory cells from multiple *pig-1(gm344) ced-3(n3692)* lines carrying the following *pig-1* transgenes: *pig-1(wt)*, the wild-type *pig-1* genomic locus (three lines, $n = 42$, 47 and 58); *pig-1(null)*, two STOP codons in the first exon (three lines, $n = 40$, 43 and 45); *pig-1(T169A)*, threonine 169 changed to alanine (three lines, $n = 40$, 41 and 48); and *pig-1(T169D)*, threonine 169 changed to aspartic acid (five lines, $n = 31$, 32, 43, 48 and 50). Error bars denote s.e.m. $P > 0.05$ (Student's *t*-test). NS, not significant ($P < 10^{-3}$). **j-l**, Merged DIC and fluorescence micrographs of *ced-3(n3692); par-4(RNAi)* (**j**), *ced-3(n3692); strd-1(RNAi)* (**k**), and *mop-25.2(ok2073); ced-3(n3692); mop-25.1(RNAi)* (**l**) larvae carrying *nIs434[P_{pgp-12::gfp}]*. Arrowheads indicate excretory and ectopic excretory-like cells. **m**, Redundant pathways mediate the elimination of ABp1pappap and other cells shed from *ced-3*-mutant embryos. Mammalian counterparts of proteins involved in cell elimination are shown in orange and blue. Scale bars, 10 μ m.

partners function in the same pathway. We believe that PIG-1 is the phosphorylation target of PAR-4 in cell shedding because inactivation of PIG-1 but of no other AMPK-related kinase caused the survival of ABp1pappap in *ced-3* mutants (Supplementary Table 7). The Tex defect of *pig-1(gm344) ced-3(n3692)* animals (89%) was higher than that of *strd-1(ok2283); ced-3(n3692)* animals (43%) (Fig. 3d and Supplementary Table 4). Thus, STRD-1 and possibly the entire PAR-4 complex are required partially for PIG-1 activation, indicating that other factors, or PIG-1 itself through autophosphorylation, also stimulate PIG-1 activity. Indeed, MELK phosphorylates itself *in vitro*¹⁸.

To investigate how PIG-1 regulates cell shedding, we reasoned that shed cells must be deficient in adhesive contacts with the embryo and explored whether *pig-1* modulates cell-adhesion complexes, specifically adherens junctions. In *C. elegans*, adherens junctions comprise a cadherin-catenin complex and a complex including DLG-1 (a homologue of the *Drosophila* discs large protein) and AJM-1. Both complexes participate in the enclosure of the embryo within a layer of epidermal cells²², a major morphogenic event that coincides with the detachment of shed cells.

We examined the expression of AJM-1, DLG-1, HMP-1 (α -catenin) and JAC-1 (p120 catenin) fused to GFP in shed cells. AJM-1, DLG-1

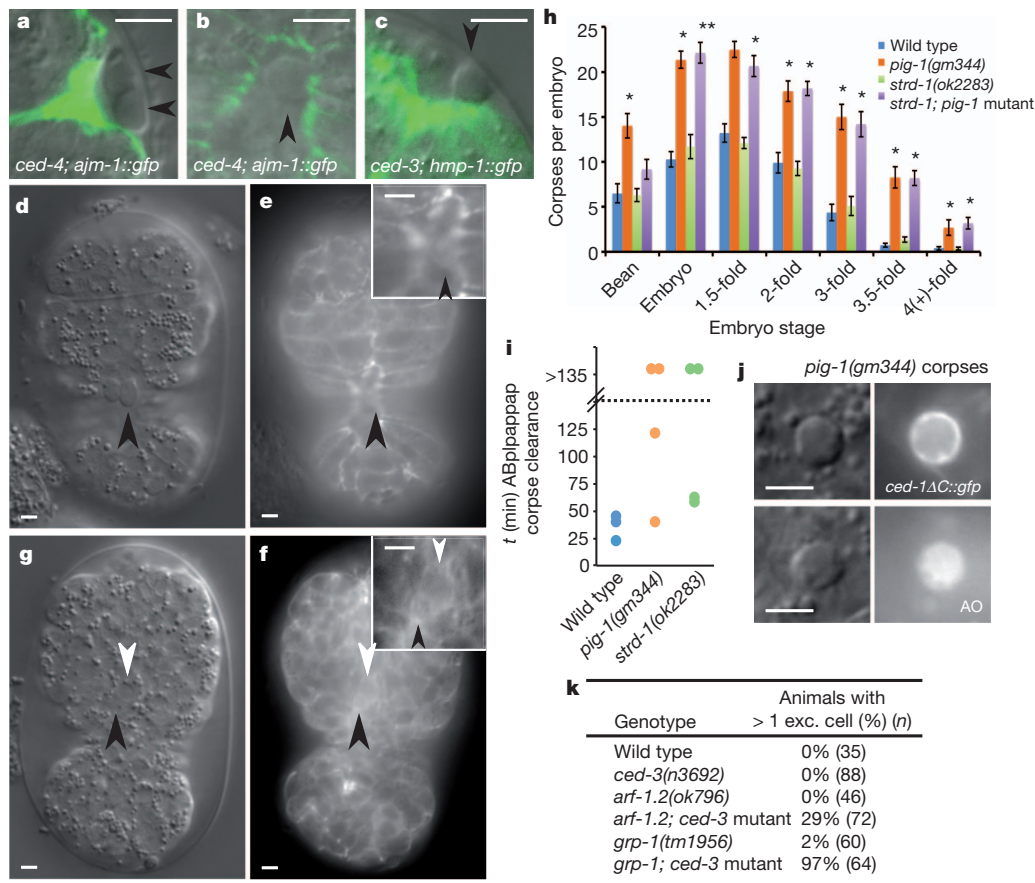


Figure 4 | Shed cells lack cell-adhesion molecules that are inappropriately expressed in *pig-1* mutants, possibly because of impaired endocytosis. **a–c**, Merged DIC and confocal fluorescence micrographs of shed cells from *ced-4(n1162); jclIs1[ajm-1::gfp]* (**a**, **b**) and *ced-3(n3692); jclIs17[hmp-1::gfp]* (**c**) eggs. Arrowheads indicate detached shed cells. **d–g** DIC and epifluorescence micrographs of *ced-3(n3692); jclIs17[hmp-1::gfp]* (**d**, **e**) and *pig-1(gm344) ced-3(n3692); jclIs17[hmp-1::gfp]* (**f**, **g**) embryos just before the completion of ventral enclosure. Black arrowheads indicate ABplpappap (**d**, **e**) or its descendant (**f**, **g**); white arrowheads indicate excretory cell (ABplpappap). Insets, magnification of ABplpappap or the excretory cell and the ABplpappap

and HMP-1 were not visible at the surface of *ced-3*- or *ced-4*-mutant shed cells that had recently detached from either the anterior sensory depression (Fig. 4a, c and Supplementary Fig. 7b) or the ventral cleft (Fig. 4b and Supplementary Fig. 7a) (JAC-1::GFP was detectable in some shed cells; Supplementary Fig. 7c). The absence of cell-adhesion proteins was noteworthy given their presence on adjacent epidermal cells (Fig. 4a–c and Supplementary Fig. 7). ABplpappap in *ced-3*-mutant embryos remained at the ventral surface of the embryo from its generation until its detachment and never expressed HMP-1::GFP (Fig. 4d, e; data not shown). By contrast, in *pig-1* *ced-3* mutant embryos, ABplpappap ingress dorsally during ventral enclosure (Fig. 4f), matching the movement of the excretory cell away from the ventral surface¹⁴. Notably, ABplpappap and its descendants exhibited a uniform expression of HMP-1::GFP at the cell surface (Fig. 4g), indicating that a loss of *pig-1* function results in the inappropriate expression of cell-adhesion molecules. Thus, PIG-1 might facilitate cell shedding by preventing the expression of cell-adhesion molecules on the surfaces of shed cells.

Several lines of evidence suggest that *pig-1* promotes shed-cell detachment through endocytosis. Many cell corpses, including the ABplpappap corpse, were not cleared efficiently in *pig-1*-mutant embryos (Fig. 4h, i). The *pig-1*-mutant cell corpses were encircled rapidly by the engulfment receptor CED-1 and stained positively with acridine orange, a marker of internalized corpses (Fig. 4j and

Supplementary Fig. 8). These results demonstrate that the delay in clearance was after engulfment and reflected a defect in corpse degradation, a process that requires endocytic pathway components²³. Interestingly, a recent study showed that two *C. elegans* ARF GTPase genes, *arf-1.2* and *arf-6*, and a gene coding for an ARF GTPase-activating protein, *cnt-2*, have functions in cell-fate determination similar to those of *pig-1* as well as roles in receptor-mediated endocytosis²⁴. Mammalian ARF GTPases function in endocytosis²⁵ and can remove cadherin complexes from the cell surface²⁶. We noted that inactivation of ARF GTPase genes *arf-1.2* or *arf-3* or the ARF guanine exchange factor gene *grp-1* also produced ectopic excretory cells in *ced-3* mutants (Fig. 4k and Supplementary Table 10).

Taken together, our observations suggest that the PAR-4 complex, PIG-1 and ARF GTPases promote the detachment of shed cells through the endocytosis-mediated removal of cell-adhesion molecules from the cell surface. Thus, the programmed elimination and apoptosis of at least eight *C. elegans* cells can be accomplished through either canonical caspase-mediated apoptosis involving the engulfment of dying cells or a caspase-independent shedding mechanism that also results in apoptosis and that requires the PAR-4 complex and the AMPK-related kinase PIG-1 (Fig. 3m). These two mechanisms are functionally redundant, as ABplpappap and the other shed cells survive only in mutants in which both pathways are disrupted.

We propose that cell shedding is an evolutionarily conserved mechanism of cell elimination. Many epithelia extrude cells constitutively to maintain tissue homeostasis, and the shed epithelial cells of vertebrates share features with the shed cells of caspase-deficient *C. elegans* embryos. First, like other shed epithelial cells⁴, shed intestinal enterocytes frequently show apoptotic markers^{27,28}, including caspase-3 activation and TUNEL reactivity. Second, despite the apoptotic appearance of shed enterocytes, the intestinal epithelia of *Casp3*^{-/-}, *Apaf1*^{-/-}, *Bax*^{-/-}, *Bak1*^{-/-} or *Bcl2*-overexpressing mice are not grossly abnormal^{3,29}, suggesting that cell extrusion is not dependent on caspase-mediated cell killing. Third, *LKB1* mutations cause Peutz–Jeghers syndrome³⁰, which is characterized by intestinal hamartomas (polyps) containing excess epithelial cells. It is possible that *LKB1* mutations contribute to polyp formation by causing a defect in the extrusion of epithelial cells. On the basis of our observations of *C. elegans*, we predict that the PIG-1 homologue MELK could be a target of *LKB1* in the gastrointestinal tract and that mutations of *MELK* might also impair enterocyte shedding and cause a polyposis phenotype in the mammalian intestine.

METHODS SUMMARY

DIC and epifluorescence micrographs were obtained using an Axioskop II (Zeiss) compound microscope, an ORCA-ER CCD camera (Hamamatsu) and OpenLab software (Agilent) and modified using ImageJ software (National Institutes of Health). For the time-lapse experiments, early (two- or four-cell stage) embryos were dissected from gravid adults, mounted on a slide with a 4% agar pad and covered with a cover slip sealed to the slide with petroleum jelly to prevent the preparation from drying. The developing embryos were imaged every 4 min for a total of 300 min, and at each time point a Z-stack of 50 images spaced at 0.6 µm was acquired. Confocal microscopy was performed using a Zeiss LSM 510 instrument, and the resulting images were modified using ImageJ software. Shed cells or floaters were counted in eggs between the 2-fold and 3.5-fold stages of development (approximately 450–600 min after the first cell division) using a ×100 objective equipped with DIC optics. The numbers of excretory cells and excretory-like cells were counted in L3 larvae carrying the *P_{gpp-12::gfp}* transcriptional reporter using a ×100 objective. A cell was scored as being ‘excretory-cell-like’ if it was located in the anterior third of the animal and its nucleus strongly expressed GFP.

Full Methods and any associated references are available in the online version of the paper.

Received 4 April 2011; accepted 17 May 2012.

Published online 15 July 2012.

1. Fuchs, Y. & Steller, H. Programmed cell death in animal development and disease. *Cell* **147**, 742–758 (2011).
2. Reddien, P. W. & Horvitz, H. R. The engulfment process of programmed cell death in *Caenorhabditis elegans*. *Annu. Rev. Cell Dev. Biol.* **20**, 193–221 (2004).
3. Yuan, J. & Kroemer, G. Alternative cell death mechanisms in development and beyond. *Genes Dev.* **24**, 2592–2602 (2010).
4. Rosenblatt, J., Raff, M. C. & Cramer, L. P. An epithelial cell destined for apoptosis signals its neighbors to extrude it by an actin- and myosin-dependent mechanism. *Curr. Biol.* **11**, 1847–1857 (2001).
5. Ellis, H. M. & Horvitz, H. R. Genetic control of programmed cell death in the nematode *C. elegans*. *Cell* **44**, 817–829 (1986).
6. Abraham, M. C., Lu, Y. & Shaham, S. A morphologically conserved nonapoptotic program promotes linker cell death in *Caenorhabditis elegans*. *Dev. Cell* **12**, 73–86 (2007).
7. Shaham, S., Reddien, P. W., Davies, B. & Horvitz, H. R. Mutational analysis of the *Caenorhabditis elegans* cell-death gene *ced-3*. *Genetics* **153**, 1655–1671 (1999).
8. Wu, Y. C., Stanfield, G. M. & Horvitz, H. R. NUC-1, a *Caenorhabditis elegans* DNase II homologue, functions in an intermediate step of DNA degradation during apoptosis. *Genes Dev.* **14**, 536–548 (2000).

9. Shaham, S. Identification of multiple *Caenorhabditis elegans* caspases and their potential roles in proteolytic cascades. *J. Biol. Chem.* **273**, 35109–35117 (1998).
10. Sulston, J. E. & Horvitz, H. R. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.* **56**, 110–156 (1977).
11. Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64–119 (1983).
12. Potts, M. B. & Cameron, S. Cell lineage and cell death: *Caenorhabditis elegans* and cancer research. *Nature Rev. Cancer* **11**, 50–58 (2011).
13. Buechner, M. Tubes and the single *C. elegans* excretory cell. *Trends Cell Biol.* **12**, 479–484 (2002).
14. Abdus-Saboor, I. et al. Notch and Ras promote sequential steps of excretory tube development in *C. elegans*. *Development* **138**, 3545–3555 (2011).
15. Cordes, S., Frank, C. A. & Garriga, G. The *C. elegans* MELK ortholog PIG-1 regulates cell size asymmetry and daughter cell fate in asymmetric neuroblast divisions. *Development* **133**, 2747–2756 (2006).
16. Ou, G., Stuurman, N., D'Ambrosio, M. & Vale, R. D. Polarized myosin produces unequal-size daughters during asymmetric cell division. *Science* **330**, 677–680 (2010).
17. Shackelford, D. B. & Shaw, R. J. The *LKB1*–AMPK pathway: metabolism and growth control in tumour suppression. *Nature Rev. Cancer* **9**, 563–575 (2009).
18. Lizcano, J. M. et al. *LKB1* is a master kinase that activates 13 kinases of the AMPK subfamily, including MARK/PAR-1. *EMBO J.* **23**, 833–843 (2004).
19. Boudeau, J. et al. MO25^{α/β} interact with STRAD^{α/β} enhancing their ability to bind, activate and localize *LKB1* in the cytoplasm. *EMBO J.* **22**, 5102–5114 (2003).
20. Zeqiraj, E., Filippi, B. M., Deak, M., Alessi, D. R. & van Aalten, D. M. F. Structure of the *LKB1*–STRAD–MO25 complex reveals an allosteric mechanism of kinase activation. *Science* **326**, 1707–1711 (2009).
21. Narbonne, P., Hyenne, V., Li, S., Labbé, J.-C. & Roy, R. Differential requirements for STRAD in *LKB1*-dependent functions in *C. elegans*. *Development* **137**, 661–670 (2010).
22. Labouesse, M. Epithelial junctions and attachments. *WormBook* **13**, 1–21 (2006).
23. Kinchen, J. M. et al. A pathway for phagosome maturation during engulfment of apoptotic cells. *Nature Cell Biol.* **10**, 556–566 (2008).
24. Singhvi, A. et al. The Arf GAP CNT-2 regulates the apoptotic fate in *C. elegans* asymmetric neuroblast divisions. *Curr. Biol.* **21**, 948–954 (2011).
25. D'Souza-Schorey, C. & Chavrier, P. ARF proteins: roles in membrane traffic and beyond. *Nature Rev. Mol. Cell Biol.* **7**, 347–358 (2006).
26. D'Souza-Schorey, C. Disassembling adherens junctions: breaking up is hard to do. *Trends Cell Biol.* **15**, 19–26 (2005).
27. Potten, C. S. Stem cells in gastrointestinal epithelium: numbers, characteristics and death. *Phil. Trans. R. Soc. Lond. B* **353**, 821–830 (1998).
28. Watson, A. J. M., Duckworth, C. A., Guan, Y. & Montrose, M. H. Mechanisms of epithelial cell shedding in the mammalian intestine and maintenance of barrier function. *Ann. NY Acad. Sci.* **1165**, 135–142 (2009).
29. Coopersmith, C. M., O'Donnell, D. & Gordon, J. I. *Bcl-2* inhibits ischemia-reperfusion-induced apoptosis in the intestinal epithelium of transgenic mice. *Am. J. Physiol.* **276**, G677–G686 (1999).
30. Hemminki, A. et al. A serine/threonine kinase gene defective in Peutz–Jeghers syndrome. *Nature* **391**, 184–187 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Z. Zhou, T. Hirose and S. Nakano for reporter constructs; B. Castor, E. Murphy and R. Droste for determining DNA sequences; S. Dennis for screening of the deletion library; N. An for strain management; and, J. Yuan, S. Nakano, N. Paquin, C. Engert and A. Saffer for discussions. The *Caenorhabditis* Genetic Center, which is funded by the National Institutes of Health National Center for Research Resources (NCRR), provided many strains. S. Mitani provided *grp-1(tm1956)*. D.P.D. was supported by post-doctoral fellowships from the Damon Runyon Cancer Research Foundation and from the Charles A. King Trust. H.R.H. is the David H. Koch Professor of Biology at the Massachusetts Institute of Technology and an Investigator at the Howard Hughes Medical Institute.

Author Contributions D.P.D. and H.R.H. designed the experiments, analysed the data and wrote the manuscript. D.P.D. and V.H. performed the experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to H.R.H. (horvitz@mit.edu).

METHODS

Strains. All *C. elegans* strains were cultured as described previously³¹ and maintained at 20 °C unless noted otherwise. We used the Bristol strain N2 as the wild-type strain. The mutations used in our experiments were as follows: LGI: *ced-1(e1735)*, *cep-1(gk138)*, *csf-3(n4872)*, *dap-1(gk219)*, *gpb-2(ad541)*, *lin-35(n745)*, *nIs433[P_{pgp-12::gfp}, unc-76(+)]*; LGII: *cad-1(j1)*, *csf-1(n4967)*, *mop-25.2(ok2073)*; LGIII: *arf-1.2(ok796)*, *ced-4(n1162)*, *ced-6(n1813)*, *ced-7(n1892)*, *ced-9(n1950)*, *n1950 n2077*, *ced-12(n3261)*, *grp-1(tm1956)*, *strd-1(ok2283)*, *nIs400[P_{ced-1::ced-1ΔC::gfp}, P_{myo-2::dsRed}]*; LGIV: *ced-2(e1752)*, *ced-3(n717)*, *n2452*, *n3692*, *ced-5(n1812)*, *ced-10(n1993)*, *csf-2(n4871)*, *ham-1(n1438)*, *lin-24(n4294)*, *lin-33(n4514)*, *pig-1(gm344)*, *n5433*, *n5437*, *jcls1[ajm-1::gfp, rol-6(su1006)]*; LGV: *unc-76(e911)*, *par-4(it57)*, *egl-1(n1084)*, *n3082*, *n3330*, *crt-1(ok948)*, *unc-68(e540)*, *unc-51(e369)*, *nIs342[P_{egl-1::gfp}, lin-15(+)]*; LGX: *nuc-1(e1392)*, *lin-15AB(n765)*, *nIs434[P_{pgp-12::gfp}, unc-76(+)]*; unknown linkage: *nIs398[P_{dyn-1::mfge8::Venus}, P_{myo-2::dsRed}]*, *jcls17[hmp-1::gfp, dlgl-1::dsRed, rol-6(su1006)]*, *nIs201[P_{unc-25::4xNLS::mStrawberry}, plin-15EK]*, *xnIs17[dlgl-1::gfp, rol-6(su1006)]*, *syIs107[lin-3::gfp]*, *jcls25[P_{hmr-1::jac-1::gfp::unc-54(3' UTR)}, rol-6(su1006)]*; extrachromosomal arrays: *nEx1747*, *nEx1748* and *nEx1749 [pig-1(wt), P_{pgp-12::gfp}, P_{myo-2::dsRed}]*; *nEx1758*, *nEx1759* and *nEx1760 [pig-1(STOP), P_{pgp-12::gfp}, P_{myo-2::dsRed}]*; *nEx1755*, *nEx1756* and *nEx1757 [pig-1(T169A), P_{pgp-12::gfp}, P_{myo-2::dsRed}]*; *nEx1831*, *nEx1832*, *nEx1833*, and *nEx1834 [pig-1(T169D), P_{pgp-12::gfp}, P_{myo-2::dsRed}]*.

Plasmids. The *P_{egl-1::gfp}* transcriptional reporter and the *P_{dyn-1::Mfge8::Venus}* (pVV59) translation reporter have been described previously^{32,33}. The *P_{unc-25::mStrawberry}* transcriptional reporter (pSN223) was constructed using PCR to amplify a 1.8 kb fragment of the *unc-25* promoter region with the primers 5'-CGAATTTTTCATGCAAAAAACACCCACTTTTTCATC-3' and 5'-CGG GATCCTCGAGCACACATCATCTTCGTCAGCAGC-3'. The resulting amplicon was digested with SphI and BamHI and ligated into pSN199, which encodes the 4xNLS::mStrawberry fusion. pSN199 was constructed by replacing the AgeI-EcoRI fragment of pPD122.56 with the AgeI-EcoRI fragment from the plasmid mStrawberry 6. The transcriptional reporter *P_{pgp-12::gfp}* (pSN359) was constructed using PCR to amplify a 2.9 kb fragment of the *pgp-12* promoter region with the primers 5'-GCAAGCTTGTGCTTGCAGTGAACCAAGAACT-3' and 5'-CCTC TAGAATTCATCAATTGGCTCGATCGCA-3'. The resulting amplicon was digested with HindIII and XbaI and ligated into pPD122.56. The *pig-1*-rescuing construct (pDD078) was constructed using PCR to amplify three overlapping genomic fragments spanning 17.5 kb with the following primer pairs: 5'-CAG TGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGGCCACATCAAA TGAAAGACG-3' and 5'-TCAGAGTTCAATATATGTTGG-3'; 5'-TGTCT ACACCACTCCAAACACC-3' and 5'-ACGACGGCATCAGATATTCG-3'; and 5'-CCAAGCGCGCAATTAACCCCTCACTAAAGGGAACAAAAGCTTG ATGATGATGTCCGTGAGC-3' and 5'-CATGTCCAACAATGGAATCG-3'. The resulting amplicons were ligated into pRS426 by homologous recombination in *Saccharomyces cerevisiae* (yeast-mediated ligation). The *pig-1(STOP)* (pDD084), *pig-1(T169A)* (pDD085) and *pig-1(T169D)* (pDD088) constructs were also generated by yeast-mediated ligation using a SfoI-NaeI-digested fragment of pDD078. *pig-1(STOP)* changes codons 8 and 9 to the stop codons TAG and TGA, respectively. *pig-1(T169A)* changes the T-loop threonine codon at position 169 to GCG, which encodes an alanine residue. *pig-1(T169D)* changes the T-loop threonine codon at position 169 to GAC, which encodes an aspartic acid residue. pDD073 (pL4440:*par-4*) was constructed using PCR to amplify the *par-4* complementary DNA with the following primers: 5'-GCGCGCGGATGG ATGCTCCGTCGACATCC-3' and 5'-GCGTCTAGACTAAGCACTATCGG TACGAG-3'. The resulting amplicon was digested with SacII and XbaI and then ligated into pL4440.

Shed cell, excretory cell, persistent corpse and extra pharyngeal cell counts. Shed cells or floaters were counted in eggs between the 2-fold and 3.5-fold stages of development (approximately 450–600 min after the first cell division) using a ×100 objective equipped with Nomarski differential interference contrast (DIC) optics. The number of excretory and excretory-like cells were counted in L3 larvae carrying the *P_{pgp-12::gfp}* transcriptional reporter using a ×100 objective. A cell was scored as an excretory cell if it was located in the anterior third of the animal and its

nucleus strongly expressed GFP. Persistent cell corpse and extra pharyngeal cell counts were performed as described previously³⁴.

Mutagenesis screen for synthetic *Tex* mutants. *ced-3(n3692)*; *nIs434* or *nIs433*; *ced-3(n3692)* L4 larvae were mutagenized with ethyl methanesulfonate as described previously³¹. Using a dissecting microscope equipped for the detection of GFP fluorescence, we screened approximately 25,000 F3 progeny of the mutagenized animals for an extra GFP+ cell near the posterior bulb of the pharynx. We showed by complementation testing and DNA sequence determination that two of the mutants isolated from this screen, *n5433* and *n5437*, were alleles of *pig-1*.

RNAi treatments. RNAi treatments were performed by feeding using RNAi constructs and reagents described previously^{35,36}. In brief, HT115 *Escherichia coli* carrying RNAi clones in the pL4440 vector were cultured overnight in Luria broth (LB) liquid media with antibiotics. Nematode growth media (NGM) plates containing 1 mM IPTG and antibiotics were seeded with the LB culture and incubated 24 h at 20 °C. Five L2 larvae were placed onto each seeded NGM plate, and the progeny of these animals were scored for either shed cells or ectopic excretory cells as described above. Control RNAi treatments were performed with the empty vector pL4440. The complete sequences of all RNAi sequences used in this study are provided in the Supplementary Information.

TUNEL and acridine orange staining. TUNEL reactions were performed as described previously^{8,37}. Fixation times (between 12 and 16 min) were optimized for labelling the floaters in *ced-5(n1812)*; *nuc-1(e1392)* eggs. *nuc-1(e1392)* facilitates the observation of TUNEL reactivity. Acridine orange staining of embryos was performed as described previously³⁸.

Transmission electron microscopy. Gravid *sem-4(n1378)*; *ced-5(n1812)* or *sem-4(n1971)*; *ced-3(n717)* adult animals were fixed by high-pressure freezing in an HPM010 high-pressure freezer by Abra. They were then substituted using an RMC FS-7500 freeze substitution system with 2% osmium, 2% water in acetone at −90 °C for 5 days, warmed to −20 °C over 14 h and then held at −20 °C for 16 h before warming to 0 °C. Samples were washed 6 × 20 min, infiltrated stepwise into Eponate 12 resin (Ted Pella) and polymerized at 60 °C. The resulting blocks were sectioned at 50 nm, post-stained with uranyl acetate and lead, and imaged with a JEM-1200EX II microscope (Jeol) using an AMT XR41 CCD camera. *sem-4* mutations cause eggs to accumulate in the hermaphrodite gonad, allowing more embryos (including more later stage embryos) to be analysed per sample.

Microscopy. Nomarski DIC and epifluorescence micrographs were obtained using an Axioskop II (Zeiss) compound microscope and OpenLab software (Agilent). Merged DIC and epifluorescence images were generated using ImageJ software (National Institutes of Health). For the time-lapse experiments, early (two- or four-cell stage) embryos were dissected from gravid adults, mounted on a slide with a 4% agar pad and covered with a cover slip that was sealed to the slide with petroleum jelly to prevent the preparation from drying. The developing embryos were imaged every 4 min for a total of 300 min, and at each time point a Z-stack of 50 images spaced at 0.6 μm was acquired. Confocal microscopy was performed using a Zeiss LSM 510 instrument, and the resulting images were prepared and modified using ImageJ software.

- Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
- Hirose, T., Galvin, B. D. & Horvitz, H. R. Six and Eya promote apoptosis through direct transcriptional activation of the proapoptotic BH3-only gene *egl-1* in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **107**, 15479–15484 (2010).
- Venegas, V. & Zhou, Z. Two alternative mechanisms that regulate the presentation of apoptotic cell engulfment signal in *Caenorhabditis elegans*. *Mol. Biol. Cell* **18**, 3180–3192 (2007).
- Schwartz, H. T. A protocol describing pharynx counts and a review of other assays of apoptotic cell death in the nematode worm *Caenorhabditis elegans*. *Nature Protocols* **2**, 705–714 (2007).
- Fraser, A. G. et al. Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408**, 325–330 (2000).
- Rual, J. F. et al. Toward improving *Caenorhabditis elegans* phenome mapping with an ORFeome-based RNAi library. *Genome Res.* **14**, 2162–2168 (2004).
- Parrish, J. Z. & Xue, D. Functional genomic analysis of apoptotic DNA degradation in *C. elegans*. *Mol. Cell* **11**, 987–996 (2003).
- Lu, Q. et al. *C. elegans* Rab GTPase 2 is required for the degradation of apoptotic cells. *Development* **135**, 1069–1080 (2008).

Site-specific DICER and DROSHA RNA products control the DNA-damage response

Sofia Francia^{1,2}, Flavia Michelini¹, Alka Saxena³, Dave Tang³, Michiel de Hoon³, Viviana Anelli^{1†}, Marina Mione^{1†}, Piero Carninci³ & Fabrizio d'Adda di Fagagna^{1,4}

Non-coding RNAs (ncRNAs) are involved in an increasingly recognized number of cellular events¹. Some ncRNAs are processed by DICER and DROSHA RNases to give rise to small double-stranded RNAs involved in RNA interference (RNAi)². The DNA-damage response (DDR) is a signalling pathway that originates from a DNA lesion and arrests cell proliferation³. So far, DICER and DROSHA RNA products have not been reported to control DDR activation. Here we show, in human, mouse and zebrafish, that DICER and DROSHA, but not downstream elements of the RNAi pathway, are necessary to activate the DDR upon exogenous DNA damage and oncogene-induced genotoxic stress, as studied by DDR foci formation and by checkpoint assays. DDR foci are sensitive to RNase A treatment, and DICER- and DROSHA-dependent RNA products are required to restore DDR foci in RNase-A-treated cells. Through RNA deep sequencing and the study of DDR activation at a single inducible DNA double-strand break, we demonstrate that DDR foci formation requires site-specific DICER- and DROSHA-dependent small RNAs, named DDRNAs, which act in a MRE11–RAD50–NBS1-complex-dependent manner (MRE11 also known as MRE11A; NBS1 also known as NBN). DDRNAs, either chemically synthesized or *in vitro* generated by DICER cleavage, are sufficient to restore the DDR in RNase-A-treated cells, also in the absence of other cellular RNAs. Our results describe an unanticipated direct role of a novel class of ncRNAs in the control of DDR activation at sites of DNA damage.

Mammalian genomes are pervasively transcribed, with most transcripts apparently not associated with coding functions^{4,5}. An increasing number of ncRNAs have been shown to have a variety of relevant cellular functions, often with very low estimated expression levels^{6–8}. DICER and DROSHA are two RNase type III enzymes that process ncRNA hairpin structures to generate small double-stranded RNAs⁹ (see Supplementary Information).

Detection of a DNA double-strand break (DSB) triggers the kinase activity of ATM, which initiates a signalling cascade by phosphorylating the histone variant H2AX (γ H2AX) at the DNA-damage site and recruiting additional DDR factors. This establishes a local self-feeding loop that leads to accumulation of upstream DDR factors in the form of cytologically detectable foci at damaged DNA sites^{3,10}. The DDR has been considered to be a signalling cascade made up exclusively of proteins, with no direct contributions from RNA species to its activation.

Oncogene-induced senescence (OIS) is a non-proliferative state characterized by a sustained DDR¹¹ and senescence-associated heterochromatic foci (SAHF)¹². Because ncRNAs participate in heterochromatin formation¹³, we investigated whether they could control SAHF and OIS. We used small interfering RNAs (siRNAs) to knockdown DICER or DROSHA in OIS cells and monitored SAHF

and cell-cycle progression. Knockdown of either DICER or DROSHA, as well as ATM as control¹⁴, restored DNA replication and entry into mitosis (Supplementary Figs 1 and 2); we did not detect overt SAHF changes, however (Supplementary Fig. 3a, b). Instead, we observed that DICER or DROSHA inactivation significantly reduced the number of cells positive for DDR foci containing 53BP1, the autophosphorylated form of ATM (pATM) and the phosphorylated substrates of ATM and ATR (pS/TQ), but not γ H2AX, without decreasing the expression of proteins involved in the DDR (Supplementary Fig. 3a–c). Importantly, the simultaneous inactivation of all three GW182-like proteins, TNRC6A, B and C, essential for the translational inhibition mediated by microRNAs (miRNAs; canonical DICER and DROSHA products involved in RNAi)¹⁵, does not affect DDR foci formation (Supplementary Fig. 4).

We next asked whether DICER or DROSHA inactivation also affects ionizing-radiation-induced DDR activation. We transiently inactivated DICER or DROSHA by siRNA in human normal fibroblasts (HNFs), exposed cells to ionizing radiation, and monitored DDR foci. We observed that a few hours after exposure to ionizing radiation, DICER or DROSHA inactivation impairs the formation of pATM, pS/TQ and MDC1, but not γ H2AX, foci without decreasing their protein levels (Fig. 1a, b and Supplementary Fig. 5). Furthermore, at an earlier time point (10 min) after ionizing radiation, 53BP1 foci were significantly reduced (Supplementary Fig. 6a). Using an RNAi-resistant form of DICER in DICER knockdown cells, we observed that re-expression of wild-type DICER, but not of a DICER endonuclease mutant (DICER44ab)¹⁶, rescues DDR foci formation (Supplementary Fig. 6b–d). The simultaneous knockdown of TNRC6A, B and C, or DICER has a comparable impact on a reporter system specific for miRNA-dependent translational repression¹⁷, but only DICER inactivation reduces DDR foci formation (Supplementary Fig. 7). To confirm further the involvement of DICER in DDR activation, we used a cell line carrying a hypomorphic allele of DICER (*DICER^{exon5}*) defective in miRNA maturation¹⁸. In *DICER^{exon5}*-irradiated cells, pATM, pS/TQ and MDC1, but not γ H2AX, foci formation is impaired without a decrease in their protein levels, and 53BP1 foci formation is delayed compared to the DICER wild-type parental cell line (Supplementary Fig. 8). These defects could be reversed by the re-expression of wild-type DICER but not of the mutant form DICER44ab (Supplementary Fig. 9). By immunoblotting, we confirmed that ATM autophosphorylation is reduced in DICER or DROSHA knockdown HNFs, and in *DICER^{exon5}* cell lines (Supplementary Fig. 10). These results indicate that DICER and DROSHA RNA products control DDR activation and act independently from canonical miRNA-mediated translational repression mechanisms.

DDR signalling enforces cell-cycle arrest at the G1/S and G2/M checkpoints³. We observed that DNA-damage-induced checkpoints were impaired in DICER- or DROSHA-inactivated cells and that

¹IFOM Foundation - FIRC Institute of Molecular Oncology Foundation, Via Adamello 16, 20139 Milan, Italy. ²Center for Genomic Science of IIT@SEMM, Istituto Italiano di Tecnologia, at the IFOM-IEO Campus, Via Adamello 16, 20139 Milan, Italy. ³Omics Science Center, RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. ⁴Istituto di Genetica Molecolare, Consiglio Nazionale delle Ricerche, Pavia 27100, Italy. [†]Present addresses: Departments of Surgery and Medicine, Weill Cornell Medical College and New York Presbyterian Hospital, 1300 York Avenue, New York, New York 10065, USA (V.A.); Institute of Toxicology and Genetics, Karlsruhe Institute of Technology, 76344 Karlsruhe, Germany (M.M.).

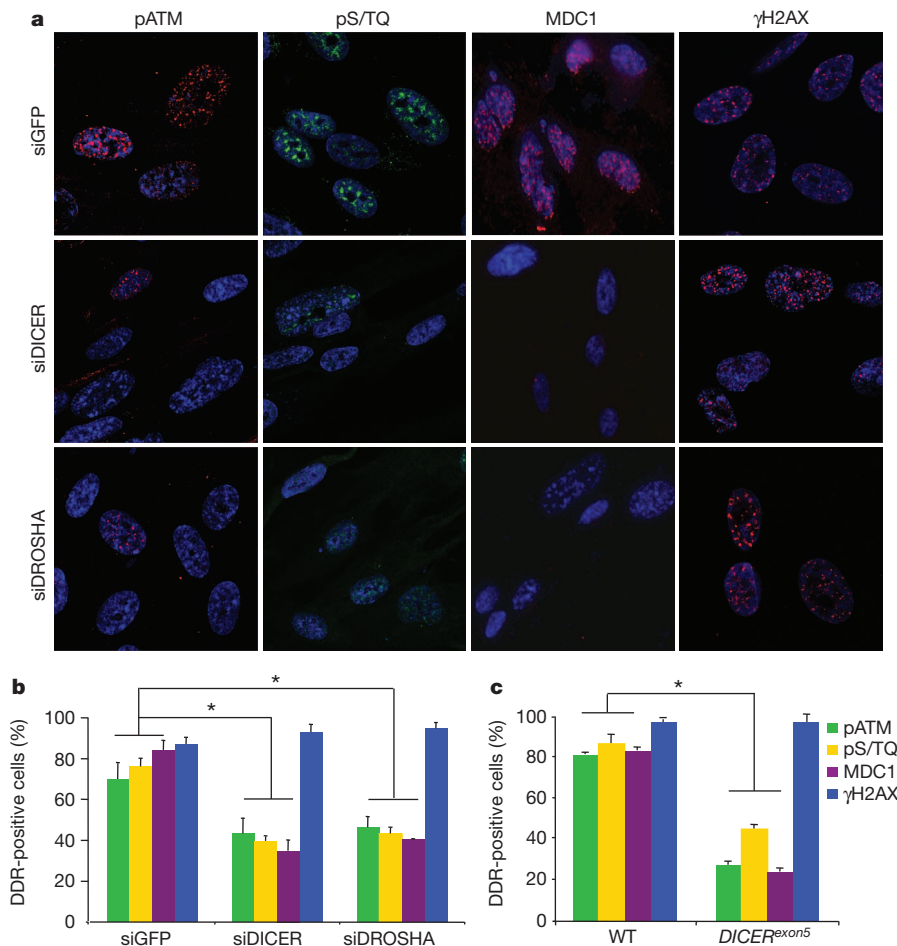


Figure 1 | DICER or DROSHA inactivation impairs DDR foci formation in irradiated cells. **a**, DICER or DROSHA knockdown WI-38 cells were irradiated (10 Gy) and fixed 7 h later. Original magnification, $\times 250$. **b**, Histogram shows the percentage of cells positive for pATM, pS/TQ, MDC1

and γ H2AX foci. **c**, Wild-type (WT) and *DICER^{exon5}* cells were irradiated (2 Gy) and fixed 2 h later. Histogram shows the percentage of cells positive for pATM, pS/TQ, MDC1 and γ H2AX foci. Error bars indicate s.e.m. ($n \geq 3$). Differences are statistically significant (* P value < 0.01).

wild-type DICER re-expression in *DICER^{exon5}* cells restores checkpoint functions whereas two independent mutant forms of DICER fail to do so (Supplementary Figs 11–13). Thus, DICER and DROSHA are required for DNA-damage-induced checkpoint enforcement.

To test the role of DICER in DDR activation in a living organism, we inactivated it by morpholino antisense oligonucleotide injection in *Danio rerio* (zebrafish) larvae¹⁹. Such Dicer inactivation results in a marked impairment of pATM and zebrafish γ H2AX accumulation in irradiated larvae as detected both by immunostaining and immunoblotting of untreated or Dicer morpholino-injected larvae and of chimaeric animals (Supplementary Figs 14 and 15).

Previous reports have shown that mammalian cells can withstand transient membrane permeabilization and RNase A treatment, enabling investigation of the contribution of RNA to heterochromatin organization and 53BP1 association to chromatin^{20,21}. We used this approach to address the direct contribution of DICER and DROSHA RNA products in DDR activation. Irradiated HeLa cells were permeabilized and treated with RNase A, leading to degradation of all RNAs, without affecting protein levels (Supplementary Fig. 16a). We observed that 53BP1, pATM, pS/TQ and MDC1 foci become markedly reduced in number and intensity upon RNA degradation whereas, similarly to DICER- or DROSHA-inactivated cells, γ H2AX is unaffected (Fig. 2a and Supplementary Fig. 16b). Notably, 53BP1, MDC1 and γ H2AX triple staining shows that RNA degradation reduces 53BP1 and MDC1 accumulation at unperturbed γ H2AX foci

(Supplementary Fig. 16c). When RNase A is inhibited, DDR foci progressively reappear within minutes and α -amanitin prevents this (Supplementary Fig. 17a, b), suggesting that DDR foci stability is RNA polymerase II dependent.

We tested whether DDR foci can reform upon addition of exogenous RNA to RNase-A-treated cells. We observed that DDR foci robustly reform in RNase-A-treated cells following their incubation with total RNA purified from the same cells, but not with transfer RNA (tRNA) control (Fig. 2b–d). Similar conclusions were reached using an inducible form of PpoI and AsiSI site-specific endonucleases^{22,23} (data not shown).

Next, we attempted to characterize the length of the RNA species involved in DDR foci reformation, which we refer to as DDRNAs. We observed that an RNA fraction enriched by chromatography for species < 200 nucleotides was sufficient to restore DDR foci (Supplementary Fig. 17c–e). To attain better size separation, we resolved total RNA on a polyacrylamide gel and recovered RNA fractions of different lengths (Supplementary Fig. 17f, g). Using equal amounts of each fraction, we observed that only the 20–35-nucleotide fraction could restore DDR foci (Fig. 2b), consistent with the size range of DICER and DROSHA RNA products.

To test the hypothesis that DDRNAs are DICER and DROSHA products, we tested DDR foci restoration with total RNA extracted from wild-type or *DICER^{exon5}* cells. Although RNA extracted from wild-type cells restores pATM, pS/TQ and 53BP1 foci, RNA from *DICER^{exon5}* cells does not (Fig. 2c, d). Importantly, RNA from

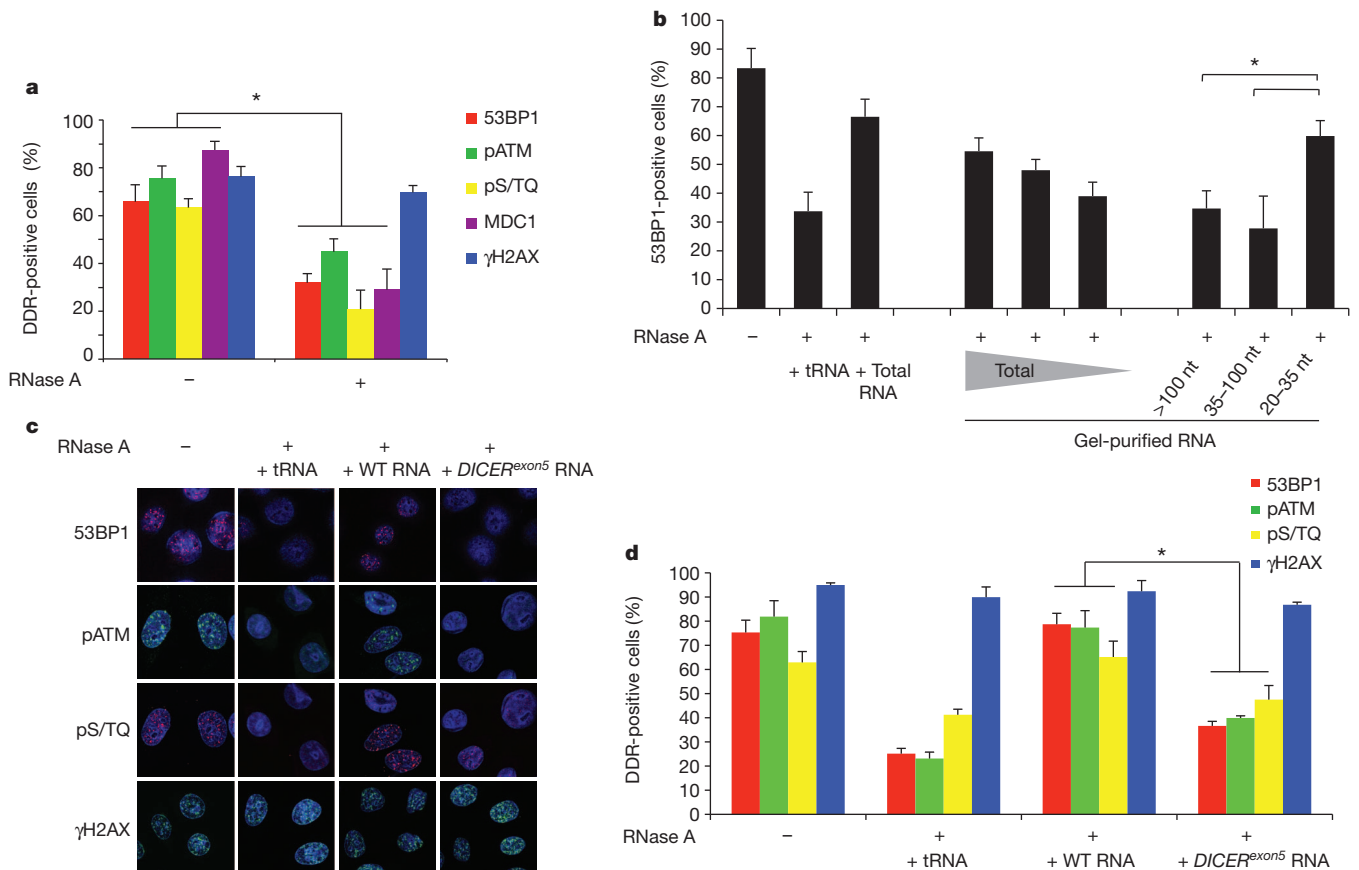


Figure 2 | Irradiation-induced DDR foci are sensitive to RNase A treatment and are restored by small and DICER-dependent RNAs. **a**, Irradiated HeLa cells (2 Gy) were treated with PBS (–) or RNase A (+) and probed for 53BP1, pATM, pS/TQ, MDC1 and γH2AX foci. Histogram shows the percentage of cells positive for DDR foci. **b**, 100, 50 or 20 ng of gel-extracted total RNA and 50 ng of RNA extracted from each gel fraction (>100, 35–100 and 20–35

nucleotides (nt)) were used for DDR foci reconstitution after RNase treatment. **c**, 53BP1, pS/TQ and pATM foci are restored in RNase-treated cells when incubated with RNA of wild-type cells but not with RNA of *DICER*^{exon5} cells or tRNA. Original magnification, ×350. **d**, Histogram shows the percentage of cells positive for DDR foci. Error bars indicate s.e.m. ($n \geq 3$). Differences are statistically significant (* P value < 0.01).

DICER^{exon5} cells re-expressing wild-type, but not endonuclease-mutant, *DICER* allows DDR foci reformation (Supplementary Fig. 18a, b). These results were reproduced using RNA extracted from cells transiently knocked down for *DICER* or *DROSHA* (Supplementary Fig. 18c–f).

Ionizing radiation induces DNA lesions that are heterogeneous in nature and random in their genomic location. To reduce this complexity, we studied a single DSB at a defined and traceable genomic locus. We therefore took advantage of NIH2/4 mouse cells carrying an integrated copy of the I-SceI restriction site flanked by arrays of Lac- or Tet-operator repeats at either sites²⁴. In this cell line, the expression of the I-SceI restriction enzyme together with the fluorescent protein Cherry-Lac-repressor allows the visualization of a site-specific DDR focus that overlaps with a focal Cherry-Lac signal (cut NIH2/4 cells). No DDR focus formation is observed overlapping with the Cherry-Lac signal in the absence of I-SceI expression (uncut NIH2/4 cells). Also in this system, RNase A treatment causes the disappearance of the 53BP1, but not the γH2AX, focus at the I-SceI-induced DSB; total RNA addition from cut cells restores 53BP1 focus formation in a dose-dependent manner (Fig. 3a, b). Therefore, a DDR focus generated on a defined DSB can disassemble and reassemble in an RNA-dependent manner.

To determine whether DDRNAs are generated at the damaged locus or elsewhere in the genome, we took advantage of the fact that the I-SceI-induced DSB is generated within an integrated exogenous sequence, which is not present in the parental cell line. As RNAs extracted from NIH2/4 or parental cells are expected to differ only in the potential presence of RNA transcripts generated at the locus, we used these two RNA preparations to attempt to restore 53BP1

focus formation at the I-SceI-induced DSB in RNase-A-treated cells. The formation of the 53BP1 focus was efficiently recovered only by RNA purified from NIH2/4 cells and not from parental cells (Fig. 3c), indicating that DDRNAs originate from the damaged genomic locus.

The MRE11–RAD50–NBS1 (MRN) complex is necessary for ATM activation²⁵, and pATM and MRE11 foci formation is sensitive to RNase A treatment in the NIH2/4 cell system (Supplementary Fig. 19a, b). To probe the molecular mechanisms by which RNA modulates DDR focus formation, we used a specific MRN inhibitor²⁶, mirin, which prevents ATM activation also in the NIH2/4 system (Supplementary Fig. 19d). In the presence of mirin, NIH2/4 RNA is unable to restore 53BP1 or pATM focus formation (Fig. 3d, e), indicating that DDRNAs act in a MRN-dependent manner.

To detect potential short RNAs originating from the integrated locus, we deep-sequenced libraries generated from short (<200 nucleotides) nuclear RNAs of cut or uncut NIH2/4 cells, as well as from parental cells expressing I-SceI as negative control. Sequencing revealed short transcripts arising from the exogenous locus (Supplementary Fig. 20a–e), 47 reads in cut cells, 20 reads in uncut cells and none in parental cells, indicating that even an exogenous integrated locus lacking mammalian transcriptional regulatory elements is transcribed and can generate small RNAs.

To test whether the identified locus-specific small RNAs are biologically active and have a causal role in DDR activation, we chemically synthesized four potential pairs among the sequences obtained and used them to attempt to restore the DDR focus in RNase-A-treated

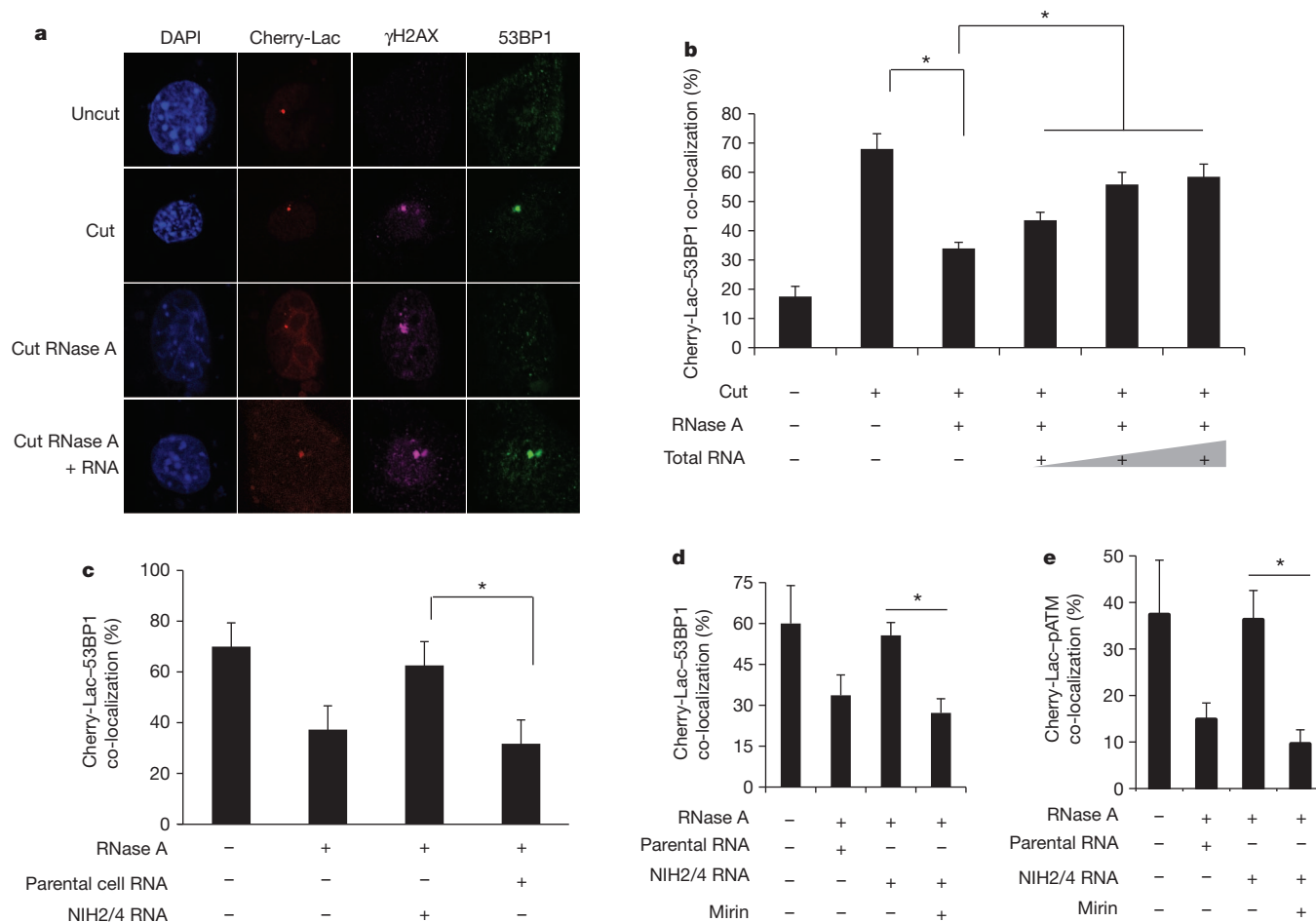


Figure 3 | Site-specific DDR focus formation is RNase A sensitive and can be restored by site-specific RNA in a MRN-dependent manner. **a**, Cut NIH2/4 cells display a 53BP1 and γ H2AX focus co-localizing with a Cherry-Lac focus. 53BP1, but not γ H2AX, focus is sensitive to RNase A and is restored by incubation with total RNA. DAPI, 4',6-diamidino-2-phenylindole. Original magnification, $\times 450$. **b**, Histogram shows the percentage of cells in which 53BP1 and Cherry-Lac foci co-localize. Addition of 50, 200 or 800 ng of RNA purified from cut NIH2/4 rescues 53BP1 foci formation in a dose-dependent

manner. **c**, RNA purified from cut NIH2/4 restores 53BP1 focus whereas RNA from parental cells expressing I-SceI does not. **d**, **e**, RNase-A-treated cut NIH2/4 cells were incubated with RNA from cut NIH2/4 cells, or parental ones, to test 53BP1 or pATM focus reformation in the presence of the MRN inhibitor mirin (100 μ M). Histogram shows the percentage of cells positive for a DDR focus. Error bars indicate s.e.m. ($n \geq 3$). Differences are statistically significant (* P value < 0.05).

cells. Notably, we observed that addition of locus-specific synthetic RNAs, but not equal amounts of control RNAs, triggers site-specific 53BP1 focus reformation over a large range of concentrations in the presence, but also in the absence, of total RNA from parental cells (Fig. 4a and Supplementary Fig. 20f). To show further the biological activity of RNAs processed by DICER, we *in vitro* transcribed both strands of the sequence spanning the locus, or a control one, and processed the resulting RNAs with recombinant DICER. *In vitro*-generated locus-specific DICER RNA products, but not control RNAs, allowed DDR focus reformation in RNase-A-treated cells even in the absence of parental RNA (Fig. 4b and Supplementary Fig. 20g, h). Overall, these results indicate that DDRNAs are small RNAs with the sequence of the damaged locus, which have a direct role in DDR activation.

To investigate the biogenesis of such RNAs *in vivo*, we performed deeper sequencing of small nuclear RNAs from cut and uncut wild-type as well as DICER or DROSHA knockdown NIH2/4 cells (Supplementary Fig. 21). As expected, DICER or DROSHA knockdown significantly reduced reads mapping to the known miRNAs (Supplementary Fig. 22). Our statistical analyses revealed that the percentage of 22–23-nucleotide RNAs arising from the locus significantly

increases in the wild-type cut sample compared to the uncut one and that DICER inactivation significantly reduces it (Supplementary Fig. 23a, b); the detectable decrease in DROSHA-inactivated cells did not reach statistical significance. Because the fraction of 22–23-nucleotide RNAs from the locus is significantly higher with respect to that of non-miRNA genomic loci, the RNAs detected are very unlikely to be random degradation products (Supplementary Fig. 23c). Finally, 22–23-nucleotide RNAs at the locus tend to have an A/U at their 5' and a G at their 3' end (Supplementary Fig. 23d), a nucleotide bias significantly different from the originating locus and from the rest of the genome.

In summary, we demonstrate that different sources of DNA damage, including oncogenic stress, ionizing radiation and site-specific endonucleases, activate the DDR in a manner dependent on DDRNAs, which are DICER- and DROSHA-dependent RNA products with the sequence of the damaged site. DDRNAs control DDR foci formation and maintenance, checkpoint enforcement and cellular senescence in cultured human and mouse cells and in different cell types in living zebrafish larvae. They act differently from canonical miRNAs, as inferred by their demonstrated biological activity independent of other RNAs and of GW182-like proteins.

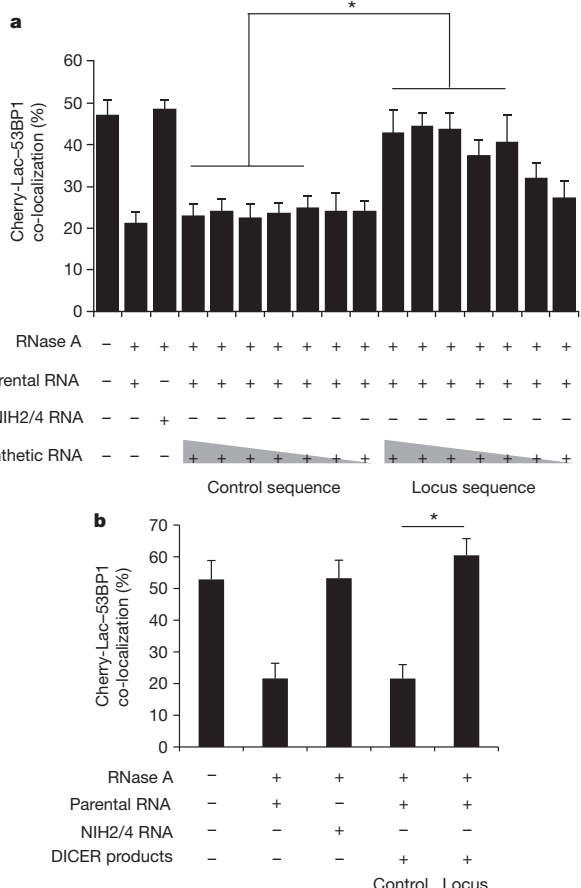


Figure 4 | Chemically synthesized small RNAs and *in vitro*-generated DICER RNA products are sufficient to restore DDR focus formation in RNase-A-treated cells in a sequence-specific manner. **a**, Chemically synthesized RNA oligonucleotides were annealed and were tested to restore DDR focus formation in RNase-A-treated cut NIH2/4 cells. Mixed with a constant amount (800 ng) of parental cell RNA, a concentration range ($1 \text{ ng } \mu\text{L}^{-1}$ to $1 \text{ fg } \mu\text{L}^{-1}$, tenfold dilution steps) of locus-specific or GFP RNAs was used. Locus-specific synthetic RNAs (down to $100 \text{ fg } \mu\text{L}^{-1}$) allow site-specific DDR activation. **b**, Small double-stranded RNAs generated by recombinant DICER were tested to restore DDR focus formation in RNase-A-treated cut NIH2/4 cells. $1 \text{ ng } \mu\text{L}^{-1}$ RNA was tested mixed with 800 ng of parental cell RNA. Locus-specific DICER RNAs, but not control RNAs, allow site-specific DDR activation. Histograms show the percentage of cells positive for DDR focus. Error bars indicate s.e.m. ($n \geq 3$). Differences are statistically significant ($*P \text{ value} < 0.05$).

METHODS SUMMARY

Details of cell cultures, plasmids, siRNAs and antibodies used, as well as descriptions of methods for immunofluorescence, immunoblotting, checkpoint assays, real-time quantitative polymerase chain reaction (PCR), zebrafish injection and transplantation, RNase A treatment, small RNA extraction and purification from gel, RNA sequencing and statistical analyses are provided in Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 8 February 2010; accepted 4 May 2012.

Published online 23 May 2012.

1. Esteller, M. Non-coding RNAs in human disease. *Nature Rev. Genet.* **12**, 861–874 (2011).
2. Krol, J., Loedige, I. & Filipowicz, W. The widespread regulation of microRNA biogenesis, function and decay. *Nature Rev. Genet.* **11**, 597–610 (2010).
3. Jackson, S. P. & Bartek, J. The DNA-damage response in human biology and disease. *Nature* **461**, 1071–1078 (2009).
4. Clark, M. B. *et al.* The reality of pervasive transcription. *PLoS Biol.* **9**, e1000625 (2011).
5. Wilusz, J. E., Sunwoo, H. & Spector, D. L. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.* **23**, 1494–1504 (2009).

6. Wang, X. *et al.* Induced ncRNAs allosterically modify RNA-binding proteins in *cis* to inhibit transcription. *Nature* **454**, 126–130 (2008).
7. Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756 (2008).
8. Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nature Rev. Genet.* **10**, 155–159 (2009).
9. Kim, V. N., Han, J. & Siomi, M. C. Biogenesis of small RNAs in animals. *Nature Rev. Mol. Cell Biol.* **10**, 126–139 (2009).
10. Lukas, J., Lukas, C. & Bartek, J. More than just a focus: the chromatin response to DNA damage and its role in genome integrity maintenance. *Nature Cell Biol.* **13**, 1161–1169 (2011).
11. d'Adda di Fagagna, F. Living on a break: cellular senescence as a DNA-damage response. *Nature Rev. Cancer* **8**, 512–522 (2008).
12. Narita, M. *et al.* Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence. *Cell* **113**, 703–716 (2003).
13. White, S. A. & Allshire, R. C. RNAi-mediated chromatin silencing in fission yeast. *Curr. Top. Microbiol. Immunol.* **320**, 157–183 (2008).
14. Di Micco, R. *et al.* Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. *Nature* **444**, 638–642 (2006).
15. Tritschler, F., Huntzinger, E. & Izaurralde, E. Role of GW182 proteins and PABPC1 in the miRNA pathway: a sense of déjà vu. *Nature Rev. Mol. Cell Biol.* **11**, 379–384 (2010).
16. Zhang, H., Kolb, F. A., Jaskiewicz, L., Westhof, E. & Filipowicz, W. Single processing center models for human Dicer and bacterial RNase III. *Cell* **118**, 57–68 (2004).
17. Nicoli, S. *et al.* MicroRNA-mediated integration of haemodynamics and Vegf signalling during angiogenesis. *Nature* **464**, 1196–1200 (2010).
18. Cummins, J. M. *et al.* The colorectal microRNAome. *Proc. Natl Acad. Sci. USA* **103**, 3687–3692 (2006).
19. Wienholds, E. *et al.* MicroRNA expression in zebrafish embryonic development. *Science* **309**, 310–311 (2005).
20. Maison, C. *et al.* Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nature Genet.* **30**, 329–334 (2002).
21. Pryde, F. *et al.* 53BP1 exchanges slowly at the sites of DNA damage and appears to require RNA for its association with chromatin. *J. Cell Sci.* **118**, 2043–2055 (2005).
22. Berkovich, E., Monnat, R. J. Jr & Kastan, M. B. Roles of ATM and NBS1 in chromatin structure modulation and DNA double-strand break repair. *Nature Cell Biol.* **9**, 683–690 (2007).
23. Iacovoni, J. S. *et al.* High-resolution profiling of γ H2AX around DNA double strand breaks in the mammalian genome. *EMBO J.* **29**, 1446–1457 (2010).
24. Soutoglou, E. *et al.* Positional stability of single double-strand breaks in mammalian cells. *Nature Cell Biol.* **9**, 675–682 (2007).
25. Stracker, T. H. & Petrini, J. H. The MRE11 complex: starting from the ends. *Nature Rev. Mol. Cell Biol.* **12**, 90–103 (2011).
26. Dupré, A. *et al.* A forward chemical genetic screen reveals an inhibitor of the Mre11–Rad50–Nbs1 complex. *Nature Chem. Biol.* **4**, 119–125 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Soutoglou, W. C. Hahn, M. Kastan, V. Orlando, R. Shiekhattar, J. Amatrudda, T. Halazonetis, E. Dejana, P. Ng and F. Nicasio for sharing reagents, M. Fumagalli and F. Rossiello for reading the manuscript, M. Dobrev, V. Matti and F. Pezzimenti for technical support, G. D'Ario for help with statistical analyses, B. Amati, M. Foiani, V. Costanzo and F.d.A.d.F. group members for help and discussions. The F.d.A.d.F. laboratory was supported by Fondazione Italiana Ricerca Sul Cancro (FIRC), Associazione Italiana Ricerca sul Cancro (AIRC) European Community's 7th Framework Programme (FP7/2007-2013) under grant agreement no. 202230, acronym "GENINCA", HFSP, AIRC, the EMBO Young Investigator Program. The initial part of this project was supported by Telethon grant no. GGP08183. P.C. was supported by 7th Framework of the European Union commission to the Dopaminet consortium, a Grant-in-Aids for Scientific Research (A) no. 20241047, Funding Program for the Next Generation World-Leading Researchers (NEXT Program) to P.C. and a Research Grant to RIKEN Omics Science Center from MEXT. S.F. is supported by Center for Genomic Science of IIT@SEMM (Scuola Europea di Medicina Molecolare) and AIRC. M.M. was supported by Cariplo (grant no. 2007-5500) and AIRC. A.S. is supported by a JSPS fellowship P09745 and grant in aid by JSPS, and D.T. is supported by the European Union 7th Framework Programme under grant agreement FP7-People-ITN-2008-238055 ("BrainTrain" project) to P.C.

Author Contributions A.S., D.T. and P.C. planned, generated and analysed the genomics data presented in Supplementary Figs 20a–e, 21, 22b and 23. M.d.H. performed statistical analysis of the genomics data. A.S. and P.C. also edited the manuscript. M.M. and V.A. generated the data presented in Supplementary Figs 14 and 15. F.M. generated the data shown in Figs 2b, 3d, e, 4b and Supplementary Figs 2b, e, 3e, 4b, 5f, g, 6b–d, 7d, 9, 13d–f, 14d, f, 17f, g, 18a, b, 19, 20g, h and 22a and generated RNA for deep sequencing; contributed to: Supplementary Figs 16a, 5d, e, 11c, d and edited the manuscript. S.F. generated the data shown in remaining figures, contributed to experimental design and edited the manuscript. F.d.A.d.F. conceived the study, designed the experiments and wrote the manuscript.

Author Information Sequence data have been deposited in the DNA Data Bank of Japan under accession code DRA000540. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.d.A.d.F. (fabrizio.dadda@fom-ieo-campus.it).

METHODS

Cultured cells. Early-passage WI-38 cells (ATCC) were grown under standard tissue culture conditions (37 °C, 5% CO₂) in MEM supplemented with 10% fetal bovine serum, 1% L-glutamine, 1% non-essential amino acids, 1% Na pyruvate. HeLa, Phoenix ecotropic and HEK293T cell lines were grown under standard tissue culture conditions (37 °C, 5% CO₂) in DMEM, supplemented with 10% fetal bovine serum, 1% glutamine, 1% penicillin/streptomycin. DICER^{exon5} colon cancer cell lines¹⁸ were cultured in McCoy's 5A medium plus 10% fetal calf serum, 1% penicillin/streptomycin. NIH2/4 cells²⁴ were grown in DMEM, supplemented with 10% fetal bovine serum, 1% glutamine, gentamicin (40 µg ml⁻¹) and hygromycin (400 µg ml⁻¹).

H-RasV12-overexpressing senescent BJ cells were generated as described previously¹⁴. BrdU incorporation assays were carried out at least 1 week after cultures had fully entered the senescent state, as determined by ceased proliferation, DDR activation and SAHF formation. Ionizing radiation was induced by a high-voltage X-ray generator tube (Faxitron X-Ray Corporation). In general, WI-38 cells were exposed to 5 Gy and transformed cells (RKO, HCT116 and HeLa) to 2 Gy for the DDR foci formation studies. 5 Gy were used for the G2/M checkpoint assays and 10 Gy for the G1/S checkpoint assays.

Cherry-Lac and I-SceI-restriction endonuclease expressing vectors were transfected by lipofectamine 2000 (Invitrogen) in a ratio of 3:1. Sixteen hours after transfection around 70% of the cells were scored positive for DDR markers in the Lac array. For generation of DICER and DROSHA knockdown, NIH2/4 cells were infected with lentiviral particles carrying pLKO.1, shDICER or shDROSHA vectors. After 48 h cells were superinfected with Adeno Empty Vector (gift from E. Dejana) or Adeno I-SceI (gift from P. Ng). Nuclei were isolated the day after the adenoviral infection.

Antibodies. Mouse anti-γH2AX, anti-H3K9me3, rabbit polyclonal anti-pH3 (Upstate Biotechnology); anti-pS/TQ (Cell Signaling Technology); anti-H2AX, anti-H3 and anti-DICER (13D6) (Abcam); rabbit polyclonal anti-53BP1 (Novus Biological); mouse monoclonal anti-53BP1 (gift from T. Halazonetis); anti-MRE11 (gift from S. P. Jackson); anti-BrdU (Becton Dickinson); rabbit polyclonal anti-MCM2 (gift from M. Melixetian); anti-MRE11 rabbit polyclonal raised against recombinant MRE11; anti-pATM (Rockland); mouse monoclonal anti-ATM and anti-MDC1 (SIGMA); anti-Lamin A/C (Santa Cruz); anti-vinculin (clone hVIN-1), anti-β-tubulin (clone AA2) and anti-Flag M2 monoclonal antibodies (Sigma).

Indirect immunofluorescence. Cells were grown on poly-D-lysinated coverslips (poly-D-lysine was used at 50 µg ml⁻¹ final concentration) and plated (15–20 × 10³ cells per cover) 1 day before staining. DDR and BrdU staining was performed as described previously¹⁴. Cells were fixed in 4% paraformaldehyde or methanol:acetone 1:1. NIH2/4 mouse cells were fixed by 4% paraformaldehyde as described previously²⁴. Images were acquired using a wide field Olympus Biosystems Microscope BX71 and the analysis or the MetaMorph software (Soft Imaging System GmbH). Comparative immunofluorescence analyses were performed in parallel with identical acquisition parameters; at least 100 cells were screened for each antigen. Cells with more than two DDR foci were scored positive. Confocal sections were obtained with a Leica TCS SP2 or AOBs confocal laser microscope by sequential scanning.

Plasmids. DICER-Flag, DICER44ab-Flag and DICER110ab-Flag were a gift from R. Shiekhattar. DICER110ab-Flag and DICER44ab-Flag double mutants carry two amino acid substitutions in the RNase III domains of DICER (Asp 1320 Ala and Asp 1709 Ala for 44ab, and Glu 1652 Ala and Glu 1813 Ala for 110ab mutant; both mutants were reported to be deficient in endonuclease activity¹⁶). pLKO.1 shDICER-expressing vector was a gift from W. C. Hahn. Short hairpin sequence for DICER is: CCGGCCACACATCTTCAAGACTTAAC CGAGTTAAGTCTTGAAGATGTGTGGTTTTTG. pRETROSUPER shp53 was as described previously¹⁴. Short hairpin sequence for p53 was: AGTAGATTAC CACTGGAGTCTT. Cherry-Lac-repressor and I-SceI-restriction endonuclease expressing vectors were gifts from E. Soutoglou²⁴. shRNA against mouse DICER- and DROSHA-expressing vectors were a gift from W. C. Hahn. shRNA for mouse DICER: CCGGGCCTCACTTGACCTGAAGTATCTCGAGATCTTCAGGTCAAGTGAGGCTTTTT. shRNA for mouse DROSHA: CCGG CTGGATATGTGCCACACTTCTCGAGAAAGTGTGGACATATTCCAGG TTTTTG.

siRNA. The DHARMACON siGENOME SMARTpool siRNA oligonucleotide sequences for human 53BP1, ATM, DICER, DROSHA were as follows. 53BP1: GAGAGCAGAUAGUCCUUUA; GGACAAGUCUCUCAGCUAU; GAUAUC AGC UAGACAAUU; GGACAGAACCCGACAGAUUU. ATM: GAAUGUU GCUUUCUGAAUU; AGACAGAAUCCCAAAUUA; UAUUACACC UGUU UGUUAG; AGGAGGAGCUUGGGCCUUU. DICER: UAAAGUAGCUGGAA UGAUG; GGAAGAGGCUGACUAUGAA; GAAUAUCGAUCCUAUGUUC; GAUCCUAUGUCAAUCUAA. DROSHA: CAACAUAGACUACACGAUU;

CCAACUCCUCGAGGAUUA; GGCCAACUGUUAUAGAAUA; GAGUAG GCUUCGUGACUUA.

The DHARMACON siGENOME siRNA sequences for human TNRC6A, B and C were as follows. GW182/TNRC6A: GAAUAGCUCUGGUCCGCUA; GCCUAAUUAUUGGUGAUUA. TNRC6B: GCACUGCCUGAUCCGAUA; GGAAUUAAGUCGUCGUCAU. TNRC6C: CUUUAACCCUGCCAAUUA; GGUAAAGUCCUCCAUUGAUG.

siRNA against human DICER 3' UTR: CCGUAAAGUUUAAACGUUU. siRNA against GFP: AACACUUGUCACUACUUUCUC. siRNA against luciferase: CAUUCUAUCCUCUAGAGGAUGdTTd; dTTGTGUAAGAUAGGAGAUCCUAC.

siRNAs were transfected by Oligofectamine or Lipofectamine RNAi Max (Invitrogen) at a final concentration of 200 nM in OIS cells and 100 nM in HNFs. In the siRNA titration experiment OIS cells were transfected in parallel with 20 nM and 200 nM siRNA oligonucleotides. For siRNA transfection with deconvolved siRNA oligonucleotides we used 50 nM for smart pools and 12.5 nM for deconvolved siRNAs.

Real-time quantitative PCR. Total RNA was isolated from cells using TRIzol (Invitrogen) or RNeasy kit (Qiagen) according to the manufacturer's instructions, and treated with DNase before reverse transcription. For small RNA isolation we used the mirVana miRNA Isolation Kit (Ambion). cDNA was generated using the Superscript II Reverse Transcriptase (Invitrogen) and used as a template in real-time quantitative PCR analysis. TaqMan MicroRNA Assays (Applied Biosystems) were used for the evaluation of mature miR-21 and rnu44 and rnu19 expression levels (assay numbers 000397, 001094 and 001003). 18S or β-actin was used as a control gene for normalization. Real-time quantitative PCR reactions were performed on an Applied Biosystems ABI Prism 7900HT Sequence Detection System or on a Roche LightCycler 480 Sequence Detection System. The reactions were prepared using SyBR Green reaction mix from Roche. Ribosomal protein P0 (RPP0) was used as a human and mouse control gene for normalization.

Primer sequences for real-time quantitative PCR. RPP0: TTCATTGTGGGAG CAGAC (forward), CAGCAGTTTCTCCAGAGC (reverse); human endogenous DICER: AGCAACACAGAGATCTCAAACATT (forward), GCAAAGCAGG GCTTTTCAT (reverse); human endogenous and overexpressed DICER: TGTTCAGGAAGACCAGGTT (forward), ACTATCCCTCAAACACTCT TGAC (reverse); human DROSHA: GGCCCGAGAGCCTTTTATAG (forward), TGACACGCTCAACTCTTCCAC (reverse); human GW182: CAGCCAGTCA GAAAGCAGTG (forward), TGTGAGTCCAGGATCTGCTACTT (reverse); mouse DICER: GCAAGGAATGGACTCTGAGC (forward), GGGGACTTCG ATATCCTCTTC (reverse); mouse DROSHA: CGTCTCTAGAAAGTCTTAC AAGAA (forward), GGCTCAGGAGCAACTGGTAA (reverse).

RNAse A treatment and RNA complementation experiments. Cells were plated on poly-D-lysinated coverslips and irradiated with 2 Gy of ionizing radiation. One hour later HeLa cells were permeabilized with 2% Tween 20 in PBS for 10 min at room temperature while I-SceI-transfected NIH2/4 cells were permeabilized in 0.5% Tween 20 in PBS for 10 min at room temperature. RNase A treatment was carried out in 1 ml of 1 mg ml⁻¹ ribonuclease A from bovine pancreas (Sigma-Aldrich catalogue no. R5503) in PBS for 25 min at room temperature. After RNase A digestion, samples were washed with PBS, treated with 80 units of RNase inhibitor (RNaseOUT Invitrogen 40 units µl⁻¹) and 20 µg ml⁻¹ of α-amanitin (Sigma) for 15 min in a total volume of 70 µl. For experiments with mirin, NIH2/4 cells were incubated at this step also with 100 µM mirin (Sigma) or DMSO for 15 min. Then, RNase-A-treated cells were incubated with total, small or gel-extracted RNA, or the same amount of tRNA, for an additional 15 min at room temperature. If using mirin, NIH2/4 cells were incubated with total RNA in the presence of 100 µM mirin or DMSO for 25 min at room temperature. Cell were then fixed with 4% paraformaldehyde or methanol:acetone 1:1.

In complementation experiments with synthetic RNA oligonucleotides, eight RNA oligonucleotides with the potential to form four pairs were chosen among the sequences that map at the integrated locus in NIH2/4 cells, obtained by deep sequencing. Synthetic RNA oligonucleotides were generated by Sigma with a monophosphate modification at the 5' end. Sequences map to different regions of the integrated locus: two pairs map to a unique sequence flanking the I-SceI restriction site, one to the Lac-operator and one to the Tet-operator repetitive sequences. Two paired RNA oligonucleotides with the sequences of GFP were used as negative control. Sequences are reported below.

Oligonucleotide 1: 5'-AUAACAAUUGUGGAAUUCGGCGC-3', oligonucleotide 2: 5'-CGAAUUCACAAUUGUUAUCC-3', oligonucleotide 3: 5'-AU UUGUGGAAUUCGGCGCCUCUAGAGUCGAGG-3', oligonucleotide 4: 5'-CC UCGACUCUAGAGGCG-3', oligonucleotide 5: 5'-AGCGGAUACAAUUU GUGGCCACAUGUGGA-3', oligonucleotide 6: 5'-UGUGGCCACAAUUG UU-3', oligonucleotide 7: 5'-ACUCCCUAUCAGUGAUAGAGAAAAGUGA

AAGU-3', oligonucleotide 8: 5'-CUUUCACUUUUCUCUAUCACUGAUAGG GAGUG-3'. GFP 1: 5'-GUUCAGCGUGUCCGGCGAGUU-3', GFP 2: 5'-CUGCCGGACACGUGAACUU-3'.

RNAs were resuspended in 60 mM KCl, 6 mM HEPES, pH 7.5, 0.2 mM MgCl₂, at the stock concentration of 12 μ M, denatured at 95 °C for 5 min and annealed for 10 min at room temperature.

DICER RNA products were generated as follows. A 550-bp DNA fragment carrying the central portion of the genomic locus studied (three Lac repeats, the I-SceI site and two Tet repeats) was flanked by T7 promoters at both ends and was used as a template for *in vitro* transcription with the TurboScript T7 transcription kit (AMSBIO). The 500-nucleotide-long RNAs obtained were purified and incubated with human recombinant DICER enzyme (AMSBIO) to generate 22–23-nucleotide RNAs. RNA products were purified, quantified and checked on gel. As a control, the same procedure was followed with a 700-bp construct containing the RFP DNA sequence. Equal amounts of DICER RNA products generated in this way were used in a complementation experiment in NIH2/4 cells following RNase A treatment.

Small RNA preparation. Total RNA was isolated from cells using TRIzol (Invitrogen) according to the manufacturer's instructions. To generate small RNA-enriched fraction and small RNA-devoid fraction we used the *mirVana* microRNA Isolation Kit (Ambion) according to the manufacturer's instructions. The *mirVana* microRNA isolation kit uses an organic extraction followed by immobilization of RNA on glass-fibre (silica-fibres) filters to purify either total RNA, or RNA enriched for small species. For total RNA extraction ethanol is added to samples, and they are passed through a filter cartridge containing a glass-fibre filter, which immobilizes the RNA. The filter is then washed a few times and the RNA is eluted with a low ionic-strength solution. To isolate RNA that is highly enriched for small RNA species, ethanol is added to bring the samples to 25% ethanol. When this lysate/ethanol mixture is passed through a glass-fibre filter, large RNAs are immobilized, and the small RNA species are collected in the filtrate. The ethanol concentration of the filtrate is then increased to 55%, and it is passed through a second glass-fibre filter where the small RNAs become immobilized. This RNA is washed a few times, and eluted in a low ionic strength solution. Using this approach consisting of two sequential filtrations with different ethanol concentrations, an RNA fraction highly enriched in RNA species ≤ 200 nucleotides can be obtained^{18,27}.

RNA extraction from gel. Total RNA samples (15 ng) were heat denatured, loaded and resolved on a 15% denaturing acrylamide gel (1 \times TBE, 7 M urea, 15% acrylamide (29:1 acryl:bis-acryl)). Gel was run for 1 h at 180 V and stained in GelRed solution. Gel slices were excised according to the RNA molecular weight marker, moved to a 2 ml clean tube, smashed and RNA was eluted in 2 ml of ammonium acetate 0.5 M, EDTA 0.1 M in RNase-free water, rocking overnight at 4 °C. Tubes were then centrifuged 5 min at top speed, the aqueous phase was recovered and RNA was precipitated and resuspended in RNase free water.

G1/S checkpoint assay. WI-38 cells were irradiated with 10 Gy and 1 h afterwards incubated with BrdU (10 μ M) for 7 h; HCT116 cells were irradiated at 2 Gy and incubated with BrdU for 2 h. Cells were fixed with 4% paraformaldehyde and probed for BrdU immunostaining. At least 100 cells per condition were analysed.

G2/M checkpoint assay. HEK 293 calcium phosphate transfected cells were irradiated with 5 Gy and allowed to respond to ionizing-radiation-induced DNA damage in a cell culture incubator for 12, 24 or 36 h. Then, at these three time points after irradiation, together with not irradiated cells, 1 $\times 10^6$ cells were collected for fluorescence activated cell sorting (FACS) analysis, fixed in 75% ethanol in PBS, 30 min on ice. Afterwards, cells were treated 12 h with 250 μ M of RNase A and incubated for at least 1 h with propidium iodide (PI). FACS profiles were obtained by the analysis of at least 5 $\times 10^5$ cells. In the complementation experiments HEK 293 cells were transfected using Lipofectamine RNAi Max (Invitrogen) and 48 h later irradiated with 5 Gy. Cells were then treated as explained above.

Immunoblotting. Cells were lysed in sample buffer and 50–100 μ g of whole cell lysate were resolved by SDS–PAGE, transferred to nitrocellulose and probed as previously described¹⁴.

For zebrafish immunoblotting protein analysis, 72 h post-fertilization (hpf) larvae were deyolked in Krebs Ringer's solution containing 1 mM EDTA, 3 mM PMSF and protease inhibitor (Roche complete protease inhibitor cocktail). Embryos were then homogenized in SDS sample buffer containing 1 mM EDTA with a pestle, boiled for 5 min and centrifuged at 13,000 r.p.m. for 1 min. Protein concentration was measured with the BCA method (Pierce) and proteins (50–900 μ g) were loaded in an SDS-12% (for γ H2AX and H3) and SDS-6% polyacrylamide gel (for pATM and ATM), transferred to a nitrocellulose membrane, and incubated with anti- γ H2AX (1:2,000, a gift from J. Amatruda²⁸), H3 (1:10,000, Abcam), pATM (1:1,000, Rockland), ATM (1:1,000, Sigma). Immunoreactive bands were detected with horseradish-peroxidase-conjugated

anti-rabbit or anti-mouse IgG and an ECL detection kit (Pierce). Protein loading was normalized to equal amounts of total ATM and H3.

Zebrafish embryo injection, cell transplantation and staining. Zebrafish embryos at the stage of 1–2 cells were injected with a morpholino against Dicer1²⁹ diluted in Danieau buffer. The morpholino oligonucleotide was injected at a concentration of 5 ng nl⁻¹, and a volume of 2 nl per embryo. To assess the efficiency of the morpholino to block miRNA maturation, we co-injected the morpholino with *in vitro* synthesized mRNA, encoding for red fluorescent protein (RFP) and carrying three binding sites for miR126 in the 3' UTR¹⁷. The oligonucleotides carrying the binding sites for miR126 used for construction of the pCS2:RFPmiR126 sensor are: 5'-GCATTATTACTACGGTACGAATAAGG CATTATTACTACGGTACGAATAAGGCATTATTACTACGGTACGA-3' and 5'-CGTAATAATGAGTGCCATGCTTATTCGCGTAATAATGAGTGCCATGCTTATTCGCGTAATAATGAGTGCCATGCT-3'. The construct was verified by sequencing and used to synthesize mRNA *in vitro* using the mMessage Kit (Ambion). mRNA encoding for RFPmiR126 sensor was injected alone or in combination with Dicer1 morpholino at a concentration of 10 pg nl⁻¹. For cell transplantation experiments, we injected donor embryos with a mixture of *dicer1* morpholino and mRNA encoding for GFP (5 pg nl⁻¹). Approximately 20 cells were transplanted from donor embryos at dome stage (5 hpf) to uninjected host at the same stage. Successfully transplanted larvae (displaying GFP+ cells) were irradiated as described below. Mature miRNAs were reverse transcribed to produce six different cDNAs for TaqMan MicroRNA assay (30 ng of total mRNA for each reaction; Applied Biosystems). Real-time PCR reactions based on TaqMan reagent chemistry were performed in duplicate on ABI PRISM 7900HT Fast Real-Time PCR System (Applied Biosystems). The level of miRNA expression was measured using C_T (threshold cycle). Fold change was calculated as 2^{-C_T}.

For immunofluorescence in zebrafish larvae, 72 hpf larvae were irradiated with 12 Gy, fixed in 2% paraformaldehyde for 2 h at room temperature. After equilibration in 10 and 15% sucrose in PBS, larvae were frozen in OCT compound on coverslips on dry ice. Sections were cut with a cryostat at a nominal thickness of 14 μ m and collected on Superfrost slides (BDH). Antisera used were zebrafish γ H2AX (gift from J. Amatruda²⁸) and pATM (Rockland). GFP fluorescence in transplanted embryos was still easily visible in fixed embryos. Images were acquired with a confocal (Leica SP2) microscope and $\times 63$ oil immersion lens.

RNA sequencing. Nuclear RNA shorter than 200 nucleotides was purified using *mirVana* microRNA Isolation Kit. RNA quality was checked on a small RNA chip (Agilent) before library preparation. For Illumina hi Seq Version3 sequencing, spike RNA was added to each RNA sample in the RNA: spike ratio of 10,000:1 before library preparation and libraries for Illumina GA IIX were prepared without spike. An improved small RNA library preparation protocol was used to prepare libraries³⁰. In brief, adenylated 3' adaptors were ligated to 3' ends of 3'-OH small RNAs using a truncated RNA ligase enzyme followed by 5' adaptor ligation to 5'-monophosphate ends using RNA ligase enzyme, ensuring specific ligation of non-degraded small RNAs. cDNA was prepared using a primer specific to the 3' adaptor in the presence of dimer eliminator and amplified for 12–15 PCR cycles using a special forward primer targeting the 5' adaptor containing additional sequence for sequencing and a reverse primer targeting the 3' adaptor. The amplified cDNA library was run on a 6% polyacrylamide gel and the 100 bp band containing cDNAs up to 33 nucleotides long was extracted using standard extraction protocols. Libraries were sequenced after quality check on a DNA high sensitivity chip (Agilent). Multiplexed barcode sequencing was performed on Illumina GA-IIX (35 bp single end reads) and Illumina Hi seq version3 (51 bp single end reads).

Statistical analyses. Results are shown as means \pm s.e.m. *P* value was calculated by Chi-squared test. Quantitative PCR with reverse transcription results are shown as means of a triplicate \pm standard deviation (s.d.) and *P* value was calculated by Student's *t*-test as indicated. *n* stands for number of independent biological experiments.

Statistical analysis of small RNA sequencing data. Statistical significance of downregulation of normalized miRNAs in DICER and DROSHA knockdown samples was calculated using the Wilcoxon signed-rank test.

The differences in the fraction of 22–23 nucleotides versus total small RNAs at the locus between the wild-type, DICER knockdown and DROSHA knockdown before and after cut were calculated by fitting a negative binomial model to the small RNAs count data and performing a likelihood ratio test, keeping the fraction of 22–23-nucleotide versus total small RNAs at the locus fixed across conditions under the null hypothesis and allowing it to vary between conditions under the alternative hypothesis.

27. Duchaine, T. F. *et al.* Functional proteomics reveals the biochemical niche of *C. elegans* DCR-1 in multiple small-RNA-mediated pathways. *Cell* **124**, 343–354 (2006).
28. Sidi, S. *et al.* Chk1 suppresses a caspase-2 apoptotic response to DNA damage that bypasses p53, Bcl-2, and caspase-3. *Cell* **133**, 864–877 (2008).

29. Wienholds, E., Koudijs, M. J., van Eeden, F. J., Cuppen, E. & Plasterk, R. H. The microRNA-producing enzyme Dicer1 is essential for zebrafish development. *Nature Genet.* **35**, 217–218 (2003).
30. Kawano, M. *et al.* Reduction of non-insert sequence reads by dimer eliminator LNA oligonucleotide for small RNA deep sequencing. *Biotechniques* **49**, 751–755 (2010).

Protein activity regulation by conformational entropy

Shiou-Ru Tzeng¹ & Charalampos G. Kalodimos¹

How the interplay between protein structure and internal dynamics regulates protein function is poorly understood. Often, ligand binding, post-translational modifications and mutations modify protein activity in a manner that is not possible to rationalize solely on the basis of structural data¹. It is likely that changes in the internal motions of proteins have a major role in regulating protein activity^{2–7}, but the nature of their contributions remains elusive, especially in quantitative terms. Here we show that changes in conformational entropy can determine whether protein–ligand interactions will occur, even among protein complexes with identical binding interfaces. We have used NMR spectroscopy to determine the changes in structure and internal dynamics that are elicited by the binding of DNA to several variants of the catabolite activator protein (CAP) that differentially populate the inactive and active DNA-binding domain states. We found that the CAP variants have markedly different affinities for DNA, despite the CAP–DNA-binding interfaces being essentially identical in the various complexes. Combined with thermodynamic data, the results show that conformational entropy changes can inhibit the binding of CAP variants that are structurally poised for optimal DNA binding or can stimulate the binding activity of CAP variants that only transiently populate the DNA-binding-domain active state. Collectively, the data show how changes in fast internal dynamics (conformational entropy) and slow internal dynamics (energetically excited conformational states) can regulate binding activity in a way that cannot be predicted on the basis of the protein's ground-state structure.

The binding of cyclic AMP to the cAMP-binding domain (CBD) of CAP allosterically induces a reorientation of the DNA-binding domains (DBDs) in CAP, thereby switching CAP from the inactive state, which does not bind to DNA, to the active state, which binds strongly to DNA⁸ (Supplementary Figs 1 and 2). To dissect the contribution of structure and internal dynamics to binding energetics, we studied a series of allosteric mutants of CAP^{9,10} in the unliganded, cNMP-liganded (cAMP- and cGMP-liganded) and DNA-liganded states (Supplementary Fig. 3a).

NMR chemical-shift analysis (Supplementary Fig. 4) of all of the CAP variants in the unliganded and nucleotide-liganded states showed that the DBD resonances fall on the line that connects the resonances of the corresponding apo form of wild-type CAP (apo-WT-CAP; the inactive DBD state) and cAMP-bound wild-type CAP (WT-CAP-cAMP₂; the active DBD state) (Figs 1a and 2a and Supplementary Fig. 3b). This observation indicates that the CAP variants are sampling a rapid equilibrium between the inactive and active DBD conformational states. The relative populations in the two DBD states can be determined from the chemical shift of the DBD resonances (Fig. 1b and Supplementary Fig. 3b) because the average chemical shift, δ , will be a weighted fraction of the population in the two states^{8,11–13}.

Relaxation dispersion analysis^{14,15} showed that, of the nine CAP variants whose chemical shift was suggestive of the unique presence of the inactive DBD state (Supplementary Fig. 3b), two variants—CAP-T127L/S128I and CAP-S62F-cAMP₂ (ref. 16)—underwent an exchange between conformations on the microsecond-to-millisecond timescale (Supplementary Fig. 5). Data fitting was indicative of a two-site exchange process for both variants, with population of the excited

state being ~7% for CAP-T127L/S128I and ~2% for CAP-S62F-cAMP₂ (Fig. 2b and Supplementary Fig. 5a, b). The difference in chemical-shift ($\Delta\omega$) values between the major and the minor conformations of CAP-T127L/S128I and CAP-S62F-cAMP₂, as determined by relaxation dispersion measurements ($\Delta\omega_{\text{disp}}$), correlates with the absolute ¹⁵N $\Delta\omega$ values of the DBD residues measured between apo-WT-CAP (inactive) and WT-CAP-cAMP₂ (active) (Supplementary Fig. 5c). Thus, the data provide strong evidence that the excited state that the DBD transiently populates in CAP-T127L/S128I and CAP-S62F-cAMP₂ (ref. 16) closely resembles the active DBD conformation.

We have thus identified eleven CAP variants that differentially populate the active DBD state, with population of this state ranging from 2% to 100% (Fig. 1a, b). In general, the affinity of a protein for a ligand is proportional to the concentration of the active binding species¹⁷. Hence, population of the active DBD state by the CAP variants is expected to correlate with the affinity of the CAP variants for DNA, with higher populations in the active state giving rise to stronger binding (Fig. 1b, open circles). However, the experimentally measured affinities (Fig. 1c) showed no correlation with population of the active DBD state (Fig. 1b, filled circles). CAP-S62F-cAMP₂ (~2% active DBD state) and CAP-T127L/S128I (~7% active DBD state) were expected to bind with a 50- and 15-fold lower affinity than WT-CAP-cAMP₂ (~100% active DBD state); however, all three of these CAP variants bound to DNA with the same affinity (Fig. 1b, c). Similarly, CAP-A144T (~50% active) was expected to bind to DNA with only a twofold lower affinity than WT-CAP-cAMP₂; however, it bound with a 50-fold lower affinity (Fig. 1b, c). Even for CAP variants that populate the active and inactive states to the same extent (for example, CAP-A144T and CAP-A144T-cGMP₂) a 75-fold difference in binding affinity for DNA was measured experimentally. Clearly, factors in addition to the structure have a major role in modulating the affinity of the protein–DNA interactions.

To determine the energetic basis of DNA binding to the CAP variants, we used isothermal titration calorimetry (ITC) (Fig. 1c and Supplementary Fig. 6). Interestingly, the various CAP proteins bound to DNA using alternative thermodynamic strategies, with some interacting with favourable enthalpy (ΔH) and others with favourable entropy (ΔS) (Fig. 1c). Given that the CAP–DNA interface is not affected by the amino acid substitutions, it is unlikely that the observed large variations in the entropy and enthalpy of binding are caused by alterations in the CAP–DNA contacts or by the amount of solvent that is removed on complex formation. Indeed, a detailed NMR analysis showed that the chemical-shift values of the DBD residues in all of the CAP–DNA complexes are essentially identical (Fig. 1d and Supplementary Fig. 7). Because the same DNA is used to form all of the complexes and because several crystal structures have shown that the CAP–DNA binding mode is invariable¹⁸, it was expected that the surface that is buried on CAP–DNA complex formation would be essentially identical for all of the CAP variants. This idea was corroborated by our observation that the contribution of solvation to the energetics of binding was similar for all of the CAP variants, as determined by measuring the thermodynamics of interaction in water and deuterated water by ITC¹⁹ (Supplementary Information). Despite solvation making a similar contribution to the binding

¹Department of Chemistry & Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, USA.

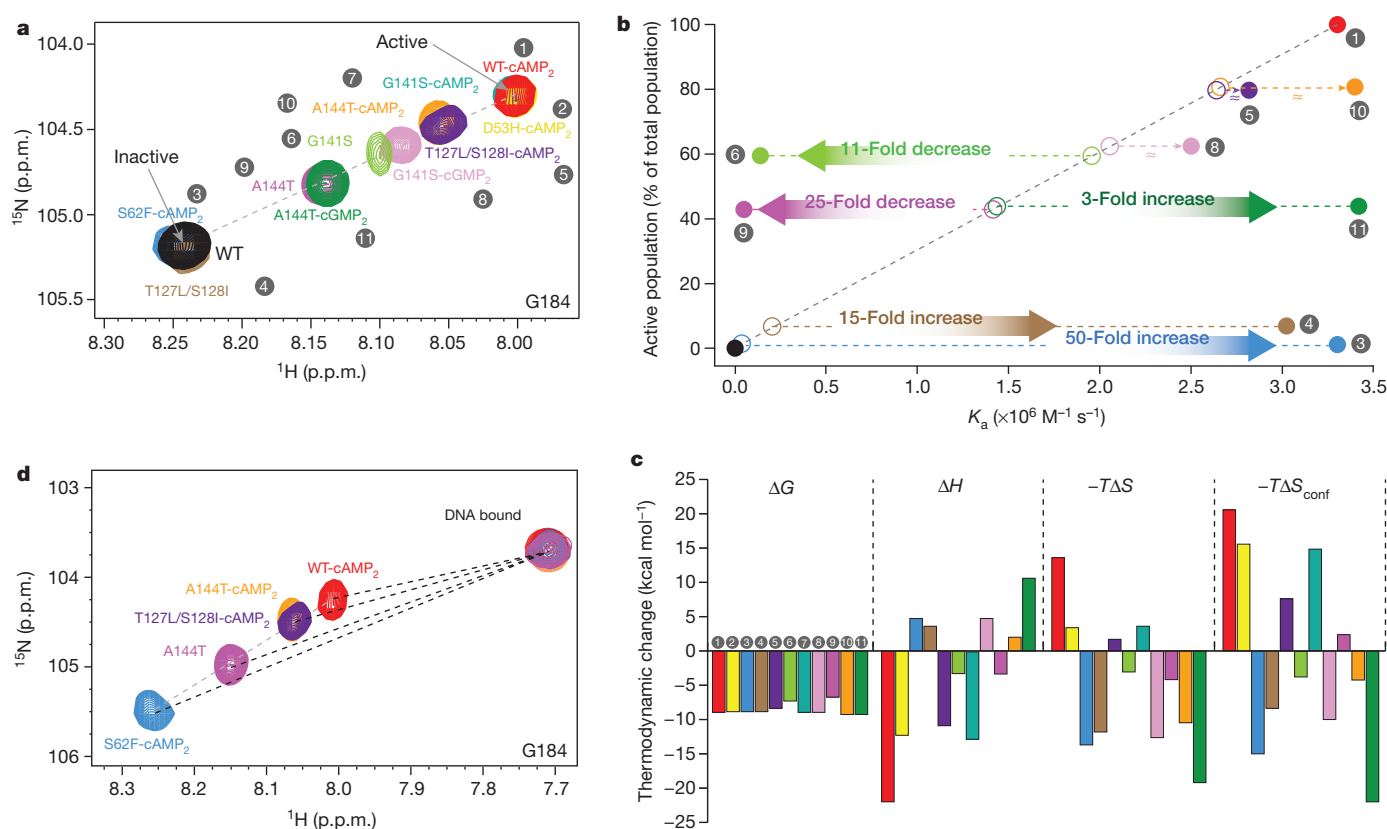


Figure 1 | CAP sampling of the active and inactive DBD states and the energetics of DNA binding. **a**, Overlaid ^1H - ^{15}N heteronuclear single quantum coherence (HSQC) spectra of the CAP variants, showing the resonance of the G184 amino acid, a residue located in the DNA recognition helix (the F helix; Supplementary Fig. 1). The chemical shift of the recognition helix residues is indicative of the conformational state of the DBD. The chemical shift of apo WT-CAP (labelled WT) indicates the inactive, DNA-binding-incompetent, conformation of the DBD, whereas the chemical shift of WT-CAP-cAMP₂ indicates the active, DNA-binding-competent, conformation of the DBD. Each CAP variant is denoted by a number (grey circles) and colour, which are the same in each figure panel. **b**, The relative population of each CAP variant in the active DBD state as a function of the variant's theoretical (open circles) and experimentally determined (filled

circles) affinity for DNA. The relative population in the active DBD state was determined from the chemical shift of the DBD resonance for those proteins interconverting on the fast exchange regime and from relaxation dispersion experiments (Supplementary Information). The theoretical affinity of CAP for DNA is the affinity expected on the basis of the population in the active DBD state considering a linear relationship. K_a , association constant. **c**, Thermodynamic components (ΔG , ΔH and $-T\Delta S$) of the binding of the CAP variants to DNA. The conformational entropy as measured by NMR is denoted $-T\Delta S_{\text{conf}}$. **d**, Overlaid ^1H - ^{15}N HSQC spectra of CAP variants in the DNA-free and DNA-bound state. The chemical shifts of the DBD resonances in the DNA-bound state are essentially identical for all of the DNA complexes of the CAP variants (see also Supplementary Fig. 7).

energetics for all of the CAP variants, the experimentally measured changes in binding entropy, $\Delta(-T\Delta S)$ (where T is temperature), spanned a remarkable range of $\sim 35 \text{ kcal mol}^{-1}$ (Fig. 1c).

To understand the origin of this large variation in binding entropy, we sought to determine the role of fast (picosecond to nanosecond) protein motions in the binding process by measuring changes in the order parameter S^2 . S^2 is a measure of the amplitude of internal motions on the picosecond-to-nanosecond timescale and can vary from $S^2 = 1$, for a bond vector with no internal motion, to $S^2 = 0$, for a bond vector that is rapidly sampling multiple orientations²⁰. We determined the changes in the S^2 of the side-chain methyl groups (S^2_{axis}) of CAP that were elicited by DNA binding (Supplementary Fig. 8). S^2 values are indicative of the amplitude of spatial fluctuations experienced by a bond vector and thus can be related to conformational entropy^{21–23}, as has previously been shown^{3,4,24–27}.

DNA binding to each of the CAP variants resulted in a notable redistribution of the amplitude of motions throughout the entire protein in a distinct manner (Supplementary Figs 9–11). DNA binding to WT-CAP-cAMP₂ resulted in the majority of the residues becoming more rigid (an increased S^2_{axis}) (Fig. 2c), thereby giving rise to a large and unfavourable change in conformational entropy (Fig. 2d). By contrast, DNA binding to either CAP-S62F-cAMP₂ or CAP-T127L/S128I resulted in the majority of the residues becoming more flexible

(Fig. 2c), thereby giving rise to a favourable conformational entropy change (-15 and -9 kcal mol^{-1} , respectively; Fig. 2d). We previously showed that the thermodynamics of the allosteric transition of the DBD from the inactive to the active state is accompanied by a small favourable entropy change¹⁶. Thus, we conclude that the calorimetrically measured large entropy change that drives the strong binding of CAP-S62F-cAMP₂ and CAP-T127L/S128I to DNA is dominated by the large and favourable conformational entropy change on complex formation (Fig. 2d). In the absence of such a significant conformational entropy change, the two CAP variants would bind poorly to DNA (Fig. 1b).

About 50% of the CAP-A144T molecules populate the active conformation (Figs 1b and 3a). Interestingly, the addition of cGMP to CAP-A144T seems to have no effect on the DBD, and CAP-A144T-cGMP₂ populates the active DBD conformation 50% of the time (Fig. 3a). Because both CAP-A144T and CAP-A144T-cGMP₂ equally populate the active species (Fig. 3a), they are expected to bind to DNA with similar affinities. However, CAP-A144T binds to DNA with a 75-fold lower affinity than does CAP-A144T-cGMP₂ (Fig. 1b, c). Thermodynamic analysis showed that the basis of this affinity difference is exclusively of an entropic nature, with the difference in entropy, $\Delta(-T\Delta S)$, for DNA binding to CAP-A144T and CAP-A144T-cGMP₂ amounting to $\sim 15 \text{ kcal mol}^{-1}$ (Figs 1c and 3c). The difference in the conformational entropy of binding between the two complexes,

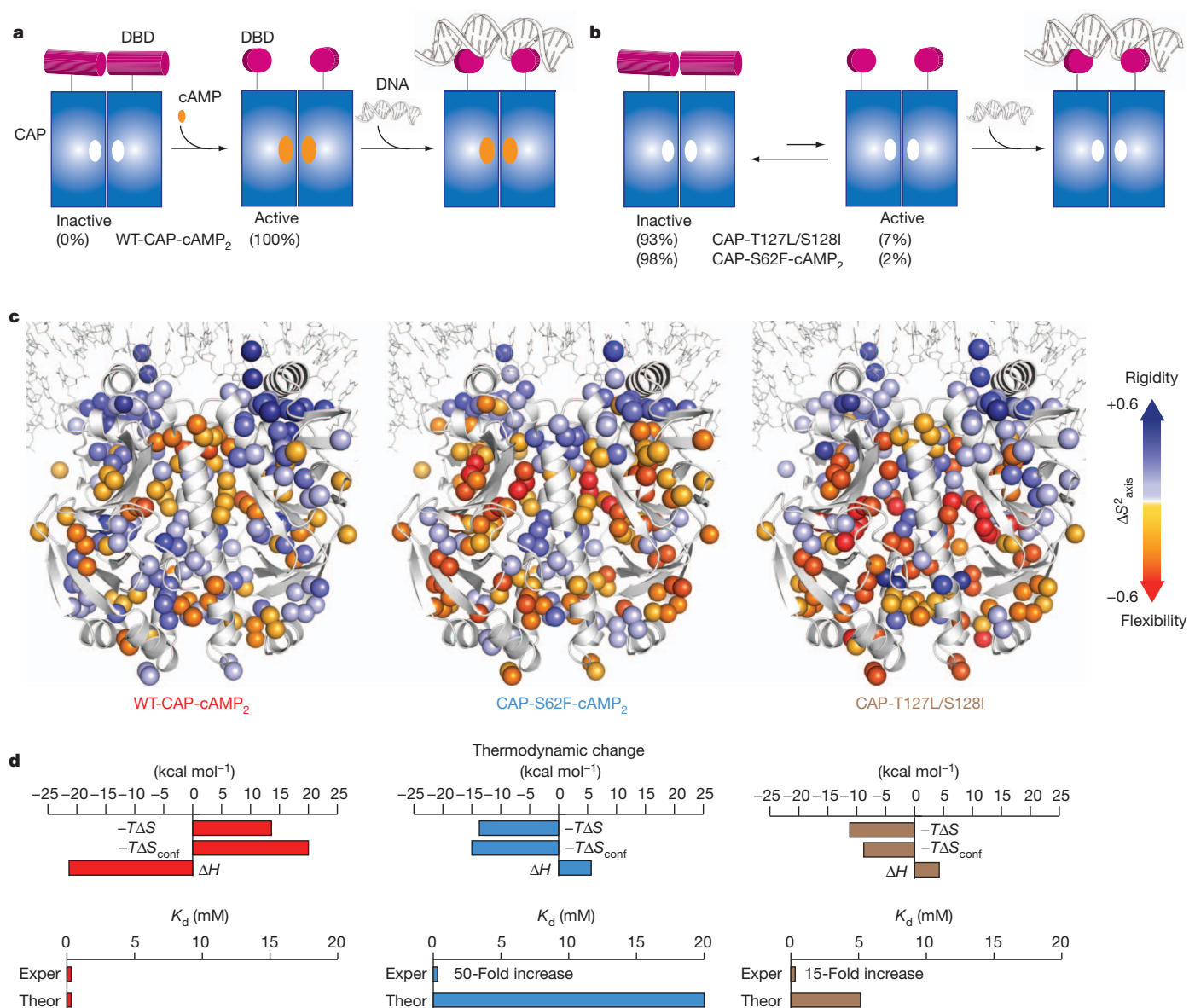


Figure 2 | Conformational entropy strongly enhances binding. **a**, The binding of cAMP to WT-CAP results in a fully (~100% populated) active DBD state. **b**, The double substitution T127L/S128I or cAMP binding to CAP-S62F results in a poorly populated active DBD state (7% and 2% of the total population, respectively). **c**, Effect of DNA binding on the methyl order parameters, S^2_{axis} , of WT-CAP-cAMP₂, CAP-S62F-cAMP₂ and CAP-T127L/S128I. The change in the order parameters is given as S^2_{axis} (after DNA binding) – S^2_{axis} (before DNA binding); therefore, positive ΔS^2_{axis} values

$\Delta(-T\Delta S_{\text{conf}})$, is ~25 kcal mol⁻¹ (Fig. 3b, c). It is of particular interest that although the two CAP variants populate the active DBD state to the same extent, they have remarkably different affinities for DNA because of the distinct responses of their conformational entropy to DNA binding, which are reflected in the internal fast motions (Fig. 3b). Similar behaviour is seen in the case of CAP-G141S and CAP-G141S-cGMP₂, with the two variants populating the active species to an almost equal extent (Fig. 1b), but with CAP-G141S binding to DNA with an order of magnitude weaker affinity, driven by the large difference in conformational entropy (Fig. 3c). Thus, our data show that unfavourable changes in conformational entropy can markedly suppress a protein's DNA binding affinity, giving rise to a binding inhibition that cannot be rationalized on the basis of protein structural data.

For some of the CAP variants, the experimentally determined affinity for DNA was similar (within twofold) to the theoretical affinity

denote enhanced rigidity of the protein on DNA binding. DNA is shown as a stick structure, and protein is shown as a ribbon structure. The methyl groups of the alanine, isoleucine, leucine, methionine and valine residues are shown as spheres. **d**, The total entropy of binding ($-T\Delta S$) as measured by ITC, the conformational entropy ($-T\Delta S_{\text{conf}}$) as determined by NMR, and the dissociation constants (K_d) (both experimentally measured (Exper) and theoretically expected (Theor) on the basis of the population in the active DBD state) are shown for each of the complexes in **c**.

(expected on the basis of the population of the active DBD state). This is the case for CAP-D53H-cAMP₂, CAP-T12L/S128I-cAMP₂, CAP-G141S-cAMP₂, and CAP-A144T-cAMP₂ (Fig. 1b, c and Supplementary Fig. 12). However, the thermodynamic strategies used by these variants to interact with DNA are distinct, with the binding being either enthalpically or entropically driven and accompanied by large changes in conformational entropy. The difference in the entropy of binding of these complexes spans a range of ~35 kcal mol⁻¹ (Fig. 1c). Dynamic analysis demonstrated that the response of each of the CAP variants to DNA binding is distinct, with the change in conformational entropy spanning a range of ~40 kcal mol⁻¹ (Fig. 1c). However, and despite the markedly different dynamic changes, there seems to be a strong enthalpy–entropy compensation that results in little or no effect on the affinity for DNA (Fig. 1b, c).

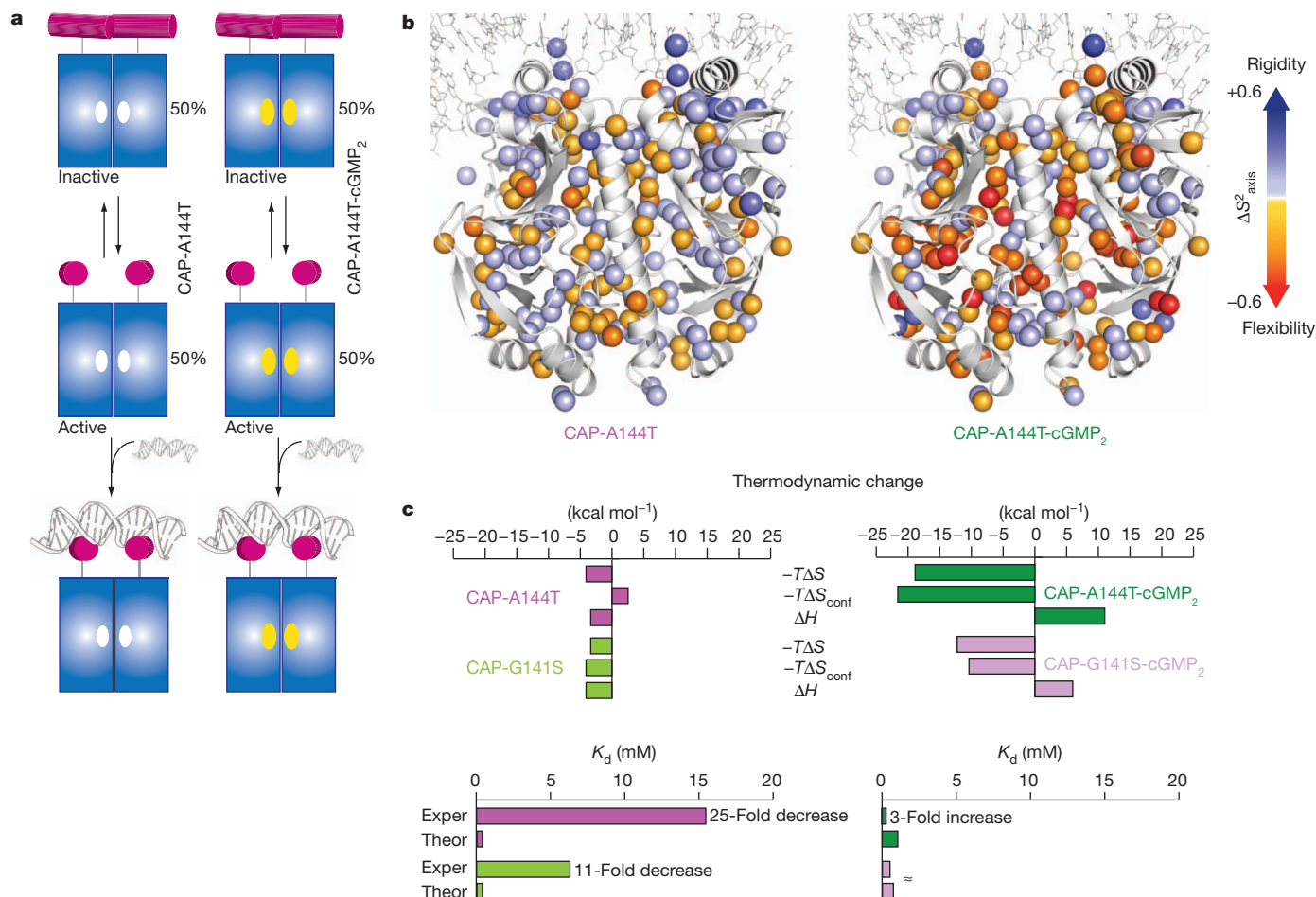
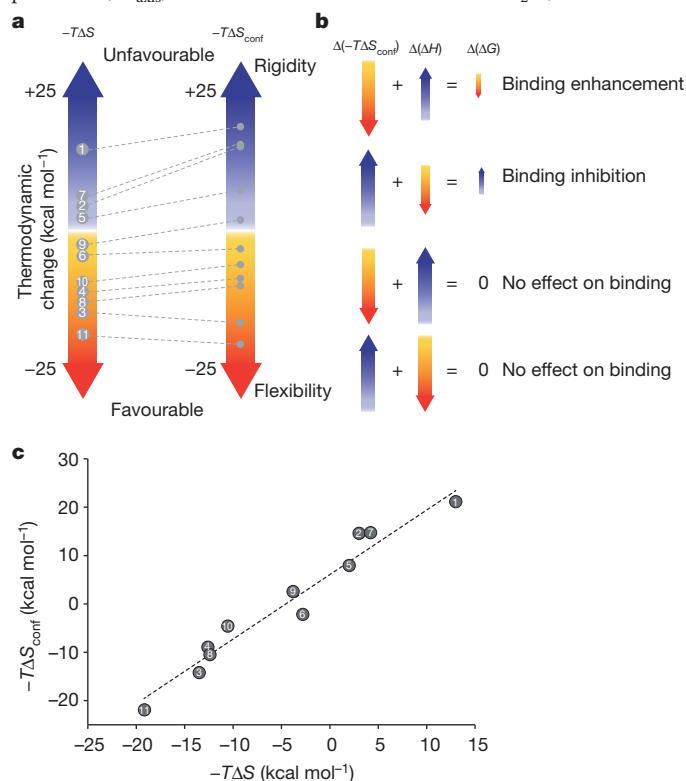


Figure 3 | Conformational entropy inhibits binding. **a**, CAP-A144T equally populates the inactive and active DBD state, and cGMP (yellow oval) binding does not perturb this equilibrium. **b**, Effect of DNA binding on the methyl order parameters, S^2_{axis} , of CAP-A144T and CAP-A144T-cGMP₂. **c**, The total

entropy of binding ($-T\Delta S$) as measured by ITC, the conformational entropy ($-T\Delta S_{conf}$) as determined by NMR, and the dissociation constants (K_d) (both experimentally measured (Exper) and theoretically expected (Theor)) are shown for two sets of complexes.



Our results show that changes in conformational entropy can dominate the total entropy of binding (Fig. 4a). Notably, there seems to be a direct relationship between the total entropy of binding and the conformational entropy of binding for the CAP complexes studied here. In fact, conformational entropy determines not only the magnitude of the total binding entropy but also whether the total entropy of binding is positive (unfavourable) or negative (favourable) (Fig. 4a, b). In some of the CAP variants, the change in conformational entropy dictates whether the enhancement or inhibition of DNA binding occurs, thereby strongly regulating protein activity (Fig. 4b). In other CAP variants, despite considerable changes in conformational entropy and thus in the total entropy of binding, the effect on the expected binding affinity is negligible (Fig. 4b). The reason for this variation is simply the anticipated enthalpy–entropy compensation²⁸, a universal phenomenon that is usually attributed to the nature of weak intermolecular interactions²⁹. In certain CAP complexes, the compensation is strong, resulting in little or no difference in the free energy of binding. However, in several complexes, the observed marked alterations in

Figure 4 | Conformational entropy dominates the total entropy of binding.

a, The total entropy of binding ($-T\Delta S$) as measured by ITC and the conformational entropy ($-T\Delta S_{conf}$) as determined by NMR for DNA binding to each of the eleven CAP variants. The numbering of the CAP variants is as in Fig. 1a. **b**, The change in conformational entropy can result in binding enhancement, binding inhibition or no effect in the expected affinity, depending on the tendency for enthalpy–entropy compensation. **c**, Relationship between the total binding entropy and the conformational entropy of binding for all eleven CAP variants.

entropy resulted in weaker enthalpy compensation, giving rise to differences in affinity of as much as two orders of magnitude (Figs 1b, c and 4b).

Notably, the relationship between the total binding entropy, as determined calorimetrically, and the conformational entropy, as assessed by changes in order parameters, is linear (Fig. 4c). There is no a priori reason for such a linear relationship. However, the observation of this linear relationship provides reassurance that two key propositions of this study are accurate. First, the entropy of binding due to all factors other than conformational entropy (that is, solvation, and translational and rotational degrees of freedom) is invariant among the CAP variants (Supplementary Information). Second, the magnitude of the error when determining the apparent conformational entropy changes by NMR is similar among the CAP variants. If either of these propositions was inaccurate, a significant deviation from linearity would be observed (Fig. 4c). Thus, our data strongly support previous observations that changes in internal dynamics can be used as a proxy for changes in conformational entropy^{24,25,30}.

In conclusion, we have shown how the interplay between a protein's fast internal motions, which are intimately related to conformational entropy, and slow internal motions, which are related to poorly populated conformational states, can regulate its activity in a way that cannot be predicted on the basis of the protein's ground-state structure. The data provide compelling evidence that the affinity of protein–ligand interactions can be strongly modulated by the global response of the internal dynamics, even when the binding interface is not perturbed. Such a marked global redistribution of the fast internal motions gives rise to substantial conformational entropy changes, with the strength of the enthalpy–entropy compensation determining the effect on the free energy of binding (Fig. 4b). This mechanism provides opportunities for regulating binding and enzymatic activities by manipulating allosteric sites without altering the binding interface.

METHODS SUMMARY

Escherichia coli BL21(DE3) cells were transformed with plasmids encoding WT-CAP or the CAP mutants and grown at 37 °C. All protein samples were purified using two chromatography steps: affinity purification on a nickel-charged Sepharose Fast Flow resin (GE Healthcare) and then size exclusion on a Superdex 200 column (GE Healthcare). NMR experiments were performed on Varian 800-MHz and 600-MHz spectrometers and Bruker 700-MHz and 600-MHz spectrometers at 32 °C. Backbone and side-chain methyl relaxation experiments and data analysis were performed as described in the Supplementary Information. ITC experiments were performed on an iTC₂₀₀ microcalorimeter (MicroCal). Full methods are available in the Supplementary Information.

Received 5 February 2012; accepted 31 May 2012.

Published online 11 July 2012.

1. Mobley, D. L. & Dill, K. A. Binding of small-molecule ligands to proteins: 'what you see' is not always 'what you get'. *Structure* **17**, 489–498 (2009).
2. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chem. Biol.* **5**, 789–796 (2009).
3. Sapienza, P. J. & Lee, A. L. Using NMR to study fast dynamics in proteins: methods and applications. *Curr. Opin. Pharmacol.* **10**, 723–730 (2010).
4. Tzeng, S.-R. & Kalodimos, C. G. Protein dynamics and allostery: an NMR view. *Curr. Opin. Struct. Biol.* **21**, 62–67 (2011).
5. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).
6. Bouvignies, G. *et al.* Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* **477**, 111–114 (2011).
7. Popovych, N., Sun, S., Ebright, R. H. & Kalodimos, C. G. Dynamically driven protein allostery. *Nature Struct. Mol. Biol.* **13**, 831–838 (2006).

8. Popovych, N., Tzeng, S.-R., Tonelli, M., Ebright, R. H. & Kalodimos, C. G. Structural basis for cAMP-mediated allosteric control of the catabolite activator protein. *Proc. Natl Acad. Sci. USA* **106**, 6927–6932 (2009).
9. Youn, H., Koh, J. & Roberts, G. Two-state allosteric modeling suggests protein equilibrium as an integral component for cyclic AMP (cAMP) specificity in the cAMP receptor protein of *Escherichia coli*. *J. Bacteriol.* **190**, 4532–4540 (2008).
10. Gekko, K., Obu, N., Li, J. & Lee, J. A linear correlation between the energetics of allosteric communication and protein flexibility in the *Escherichia coli* cyclic AMP receptor protein revealed by mutation-induced changes in compressibility and amide hydrogen–deuterium exchange. *Biochemistry* **43**, 3844–3852 (2004).
11. Volkman, B., Lipson, D., Wemmer, D. & Kern, D. Two-state allosteric behavior in a single-domain signaling protein. *Science* **291**, 2429–2433 (2001).
12. Yao, X., Rosen, M. K. & Gardner, K. H. Estimation of the available free energy in a LOV2-*Jx* photoswitch. *Nature Chem. Biol.* **4**, 491–497 (2008).
13. Li, P., Martins, I., Amarasinghe, G. & Rosen, M. Internal dynamics control activation and activity of the autoinhibited Vav DH domain. *Nature Struct. Mol. Biol.* **15**, 613–618 (2008).
14. Palmer, A. G., Grey, M. J. & Wang, C. Solution NMR spin relaxation methods for characterizing chemical exchange in high-molecular-weight systems. *Methods Enzymol.* **394**, 430–465 (2005).
15. Baldwin, A. J. & Kay, L. E. NMR spectroscopy brings invisible protein states into focus. *Nature Chem. Biol.* **5**, 808–814 (2009).
16. Tzeng, S.-R. & Kalodimos, C. G. Dynamic activation of an allosteric regulatory protein. *Nature* **462**, 368–372 (2009).
17. Bosshard, H. Molecular recognition by induced fit: how fit is the concept? *News Physiol. Sci.* **16**, 171–173 (2001).
18. Lawson, C. L. *et al.* Catabolite activator protein: DNA binding and transcription activation. *Curr. Opin. Struct. Biol.* **14**, 10–20 (2004).
19. Chervenak, M. & Toone, E. A direct measure of the contribution of solvent reorganization to the enthalpy of ligand-binding. *J. Am. Chem. Soc.* **116**, 10533–10539 (1994).
20. Ishima, R. & Torchia, D. Protein dynamics from NMR. *Nature Struct. Biol.* **7**, 740–743 (2000).
21. Akke, M., Bruschweiler, R. & Palmer, A. G. I. NMR order parameters and free energy: an analytical approach and its application to cooperative Ca²⁺ binding by calbindin D9k. *J. Am. Chem. Soc.* **115**, 9832–9833 (1993).
22. Yang, D. & Kay, L. Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. *J. Mol. Biol.* **263**, 369–382 (1996).
23. Li, Z., Raychaudhuri, S. & Wand, A. J. Insights into the local residual entropy of proteins provided by NMR relaxation. *Protein Sci.* **5**, 2647–2650 (1996).
24. Frederick, K. K., Marlow, M. S., Valentine, K. G. & Wand, A. J. Conformational entropy in molecular recognition by proteins. *Nature* **448**, 325–329 (2007).
25. Marlow, M. S., Dogan, J., Frederick, K. K., Valentine, K. G. & Wand, A. J. The role of conformational entropy in molecular recognition by calmodulin. *Nature Chem. Biol.* **6**, 352–358 (2010).
26. Petit, C. M., Zhang, J., Sapienza, P. J., Fuentes, E. J. & Lee, A. L. Hidden dynamic allostery in a PDZ domain. *Proc. Natl Acad. Sci. USA* **106**, 18249–18254 (2009).
27. Akke, M. Conformational dynamics and thermodynamics of protein–ligand binding studied by NMR relaxation. *Biochem. Soc. Trans.* **40**, 419–423 (2012).
28. Olsson, T. S. G., Ladbury, J. E., Pitt, W. R. & Williams, M. A. Extent of enthalpy–entropy compensation in protein–ligand interactions. *Protein Sci.* **20**, 1607–1618 (2011).
29. Dunitz, J. Win some, lose some: enthalpy–entropy compensation in weak intermolecular interactions. *Chem. Biol.* **2**, 709–712 (1995).
30. Wand, A. J. Dynamic activation of protein function: a view emerging from NMR spectroscopy. *Nature Struct. Biol.* **8**, 926–931 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to L. Kay for providing the methyl relaxation pulse sequences, D. Korzhnev for his help with the relaxation dispersion data analysis, and S. Kim and M.-T. Pai for their help with protein preparation and NMR experiments. This work was supported by National Science Foundation (NSF) grant MCB1121896 to C.G.K.

Author Contributions C.G.K. conceived the project. S.-R.T. and C.G.K. designed the experiments. S.-R.T. performed all of the experiments. S.-R.T. and C.G.K. analysed and interpreted the data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.G.K. (babis@rutgers.edu).

CAREERS

TURNING POINT Political criticism prompts entomologist to seek crowd-funding p.243

CAREERS BLOG The latest discussions and scientific-careers issues go.nature.com/z8g4a7

NATUREJOBS For the latest career listings and advice www.naturejobs.com



JOB APPLICATIONS

Straight to the top of the pile

The distinctions between a lengthy, technical CV and a snappy résumé can make a big difference in a job search.

BY KAREN KAPLAN

Competition for research positions is fierce, and first impressions are crucial. In most cases, a written job application represents a potential employer's first exposure to an applicant. But assembling a list of qualifications and achievements is no simple matter. Should it be a short, snappy résumé or an exhaustive CV? How should it look? How technical should it be?

Early-career researchers need to know and

understand the expectations of employers in different countries and sectors. The content, format, design and language of the application materials can influence a candidate's chances of nabbing an interview. The wrong type of document, or one that has too much, too little or irrelevant information, can propel an application to the bottom of the pile — or disqualify it altogether. Given the state of the global economy, applicants can't afford to commit faux pas that could jeopardize their chances.

In all countries, the CV and résumé provide

an overview of an applicant's experience and achievements, and candidates need to know what to include, how to format the document and how to organize and present it visually. But accepted practices differ. In the United States and Canada, a CV is comprehensive, whereas a résumé is concise. In the United Kingdom and the European Union (EU), 'CV' and 'résumé' are generally synonymous, but there are variations. Researchers need to know when one, the other or a hybrid is most suitable.

NATIONAL DIFFERENCES

A US-style CV is usually an exhaustive, chronological timeline of the applicant's education, career and accomplishments. It starts with educational history and research experience, and includes a full record of awards, publications, grants received, conferences presented at, peer-review experience, committee work and other activities. Most early-career scientists have CVs longer than ten pages, and not all the content will be relevant to the position sought.

A US-style résumé, by contrast, is a succinct and tailored summary of the applicant's work and education, which focuses on his or her qualifications for the position. It should grab the reader's attention: according to a study published this year, US recruiters spend an average of just six seconds reading a résumé before deciding whether to pass it on or bin it (W. Evans, *Eye Tracking Online Metacognition: Cognitive Complexity and Recruiter Decision Making*; TheLadders, 2012). Presentation and structure are therefore paramount: a résumé's first page must highlight the applicant's relevant skills, knowledge and capabilities.

The same study found that recruiters spend most of those six seconds on the applicant's name, educational history, current position and most recent past position. The résumé should therefore keep descriptions of pertinent career and training experience and significant grants, publications and awards very succinct. Most résumés, even for mid-career scientists, should not exceed three pages. "The résumé is typically shorter and simpler than a CV because a non-scientist is often the first reader," says Joe Tringali, managing director of Tringali and Associates, a recruitment consultancy in Manchester, New Hampshire. "It's slightly salesy — it's a marketing document, with less emphasis on publications and more on techniques and skills."

It is important for international researchers seeking US positions to note that résumés ►

K. REINICK/SHUTTERSTOCK

► should not include personal information or a personal photograph.

Appearance and layout are crucial for US résumés, although they are relatively insignificant for CVs in comparison with content. In a résumé, the applicant's name and contact information should be in large type at the top of the first page. There should be wide margins at the top, bottom and sides of each page and separating each entry; the main text must be no smaller than 11 point; and the font should be clean and easy to read.

THE RIGHT TOOLS FOR THE JOB

In general, say US hiring managers and careers advisers, applicants should send a full-length CV for research and grant applications (including fellowships) in academia, government and non-profit organizations, whereas a résumé is best for non-research posts in government, the non-profit sector and industry. For academic research posts that involve some teaching, such as at small liberal-arts colleges, or for academic or industrial postdoctoral research, the best option is a hybrid style that includes relevant skills, capabilities and knowledge on the first page. It should also contain information such as full or partial listings of grants, awards, publications and experience as a reviewer or on committees, and could total between three and eight pages. If there is a great deal more information on publications, posters, conference presentations, books authored and so on that is relevant to the position being sought, applicants can send the hybrid but append a note saying that a full CV is available.

The hybrid can also work well for professional-networking websites such as LinkedIn, where work and educational history should be informative and career experience clear, but an exhaustive list of publications, presentations, committee work and other experience is unnecessary. It is important, however, to include relevant keywords online, because many recruiters search for job candidates on networking sites and use keywords to refine and narrow their search. Careers advisers say that here, too, it is often best to link to the full CV, or add a note saying that it is available on request.

For US industrial-research applications, a résumé is almost obligatory — unless the applicant knows for certain that they are sending their submission directly to a scientist. If an advert requests a CV, résumé or both, it is best



Applicants need to be aware of cultural differences in different countries.

Barbara Janssens



Studies show that recruiters focus on CV and résumé hotspots (red) such as name and most recent job.

to send a résumé, but indicate that a full CV is available. If the advert provides no instruction and gives a generic e-mail address, such as careers@xyzbio.com, a non-scientist screener will usually view the document first, and will expect a résumé.

In the United Kingdom and the EU, the long-form CV is rare and employers in all sectors are used to a style more or less identical to the US hybrid version. Applicants to UK positions should lead with contact information and personal data such as nationality, visa status, age and whether they have a driving licence, says Sarah Blackford, head of education and public affairs at the Society for Experimental Biology in London. She notes, however, that age is optional, and there should never be a photograph. Academic CVs should include references to attached appendices in which the applicant can more fully explain their research experience, publications, grants and conference presentations.

The European Commission is trying to standardize job-application formats using an online tool for uploading CVs (go.nature.com/yxm9r1). Unlike in the United Kingdom, CVs in the EU should include a photograph, says Emilia Daniłowicz-Luebert, an immunology PhD student at the Humboldt University in Berlin and team leader for the Careers in Life Sciences Project of the Young European Bio-tech Network in Bertinoro, Italy.

In China, CV formats are similar to the long-form US style, says Zhou Zhonghe, director of the Institute of Vertebrate Paleontology and Paleoanthropology in Beijing. He says that most employers offer an online tool for uploading.

Applicants can ensure that they are prepared for any eventuality by keeping multiple versions of résumés and CVs. Matt Hepworth, a UK-born immunologist at the Humboldt

University who is seeking a second postdoc position, has two versions of a hybrid, which he has sent out for industrial and academic jobs in the United States, the EU and the United Kingdom.

The first page of both versions includes a summary of his research focus and its implications for disease, but the industrial version also has a section highlighting his relevant skills and the techniques and protocols that he has mastered. The academic version lists conference presentations, travel grants and journals for which he is a peer reviewer. Hepworth's academic version is five pages long; the industrial version is three-and-a-half. "I really tried to restrict the length, even on the academic one," says Hepworth. Staying concise seems to have paid off — Hepworth is negotiating for a US postdoc job that he hopes will eventually become a tenure-track research post.

TAILOR TO FIT

Applications should be tailored for the post of interest. When the US long-form CV is called for, candidates should tailor through their cover letter; otherwise, the CV itself should be adjusted. UK industrial-research applicants should consider the description of the position in the advert when composing their research summaries, highlights and all text outlining current and past research positions and fellowships. "Use your positions as headlines and then demonstrate through your descriptions that you can meet the needs of that business," says Alison Mitchell, deputy director of Vitae, a research-career-support organization in Cambridge, UK.

In all countries, it is likely that the first viewer of an industrial application will not be a scientist, so an applicant's résumé (in the United States) or CV (in the United Kingdom and EU) should be peppered liberally

with words and phrases mined from the advert. A computer will almost certainly do the initial screen to weed out non-viable candidates and assign ratings; applicants shouldn't risk their submission being deleted just because it didn't contain the appropriate keywords. "Give them every reason to screen you in instead of screening you out," says Tringali, who adds that résumés and CVs for industry should not be overly technical.

In the interest of brevity, industrial applications should not include appendices, says Blackford. Terry Jones, senior careers consultant at the Careers Group, University of London, agrees. "Some applicants think that hiring managers will be happier with a much longer account. But people are busy," he stresses. "You need to get over some key points pretty quickly. It's about clarity, not about endless detail." Including irrelevant information about publications, grants, awards and presentations could also send the wrong message: "If industry sees someone with a huge publications appendix, they may think this person is still hanging on to academic culture," says Blackford.

Those applying to positions in Germany, Austria and parts of Switzerland need to be aware of cultural differences, says Barbara Janssens, PhD career manager for the German Cancer Research Center in Heidelberg. If an advert is in German, employers in those countries expect CVs in the same language — not in English. CVs should include a professional photograph of the applicant, she says, and must be signed and dated. They should also include personal details such as date of birth and marital status, and copies or PDFs of diplomas and certificates.

When applying for scientific positions in most other European nations, it is safe to send application materials in English, unless the advert is in another language. In that case, applicants should contact the employer to learn what language to use.

Ultimately, say careers advisers, applicants need to suss out potential employers' expectations for format, language and other uncertainties by reading the advert, checking with mentors, reaching out to contacts who work for the employer and asking the employer themselves. The most brilliant research accomplishments can't work in an applicant's favour if the CV or résumé goes unread, as Sharon Milgram, director of the Office of Intramural Training and Education at the US National Institutes of Health in Bethesda, Maryland, points out. "Don't blow your chances," she says, "by not giving me what I want." ■

Karen Kaplan is assistant Careers editor for Nature.

TURNING POINT

Brian Fisher

Brian Fisher, an entomologist and curator at the California Academy of Sciences in San Francisco, has maintained funding from the US National Science Foundation (NSF) since 1994 to collect and study ants from around the world. He has identified more than 1,000 species and studied their evolution. After a few US politicians suggested that his grants are wasteful government spending, he started considering different ways to fund his research. This year, Fisher found some success with crowd-funding through a website called Petridish.



How did your research become politicized?

It started in 2010, when Senator John McCain (Republican, Arizona) listed my field work in East Africa — collecting ants and sharing their photos and information on the AntWeb website — as number six of what he considered the top 100 most wasteful projects funded by the American Recovery and Reinvestment Act of 2009. Since then, the project has been cited in at least six different Republican campaign commercials as an example of how President Barack Obama's administration wastes money. It has been interesting to get raked over the coals. But what caught me most off guard was that the critics, such as conservative radio personalities, weren't necessarily focused on funding for ants; they were questioning whether the government should fund science at all.

Has there been any fallout for your career?

I have a big research project, with 25 people involved worldwide. I'm worried about the next NSF funding cycle, and the negative publicity doesn't help. So I've been looking at alternative sources of funding, including crowd-funding: small contributions from online donors.

How has science funding changed since you got your PhD?

Scientists have become more like entrepreneurs, having to seek many sources of support. Ant research has always been on the fringe, getting by on crumbs of funding. But you used to be able to sustain your career on NSF funding. Now scientists need a portfolio of options.

How did you learn to create such a portfolio?

I dropped out of my first graduate programme, in biology, because the only money available required me to work on a project that I wasn't passionate about. I spent a year incubating ideas and writing grant applications — figuring out how to raise my own money — so that I could pursue research on ant diversity. Since then, I have raised well over US\$750,000 from

unconventional sources, including private donors, corporations and foundations, to create the Madagascar Biodiversity Center in Antananarivo, which identifies land for conservation and catalogues local species. I've also been able to create Ant Course: a field course offered in different countries to teach students about ant taxonomy and field-research techniques.

Describe your experience on Petridish.

I wanted to secure at least \$10,000 in funding to visit a remote, rugged, pristine forest in northwestern Madagascar, to collect ant species before the habitat is converted for cattle raising. My project was posted online for 45 days and I landed 94 backers — ranging from one who pledged more than \$5,000 to 48 who pledged about \$20 each. This was my first experience with crowd-funding and it was really hard, especially shooting the requisite video pitch. I'm used to investing three weeks of blood and sweat writing an NSF grant application, but speaking directly to the public was very different.

How might crowd-funding help science?

Scientists need alternative sources of money now, but that is just one of the benefits. Crowd-sourcing helps to democratize science — the websites let amateur scientists participate. There is a public-relations aspect — you have to make clear the relevance of your research. I think every graduate student should try to get funding in this way, because the emphasis is on communication. They would need to focus their questions and make a pitch, but a few thousand dollars could be enough to support them. Graduate students need to learn how to advocate for their field — you can't just hide inside the ivory tower. The walls are gone. ■

INTERVIEW BY VIRGINIA GEWIN